

# Exploring Social Recommendations with Visual Diversity-Promoting Interfaces

CHUN-HUA TSAI, University of Pittsburgh

PETER BRUSILOVSKY, University of Pittsburgh

The beyond-relevance objectives of recommender systems have been drawing more and more attention. For example, a diversity-enhanced interface has been shown to associate positively with overall levels of user satisfaction. However, little is known about how users adopt diversity-enhanced interfaces to accomplish various real-world tasks. In this paper, we present two attempts at creating a visual diversity-enhanced interface that presents recommendations beyond a simple ranked list. Our goal was to design a recommender system interface to help users explore the different relevance prospects of recommended items in parallel and to stress their diversity. Two within-subject user studies in the context of social recommendation at academic conferences were conducted to compare our visual interfaces. Results from our user study show that the visual interfaces significantly reduced the exploration efforts required for given tasks and helped users to perceive the recommendation diversity. We show that the users examined a diverse set of recommended items while experiencing an improvement in overall user satisfaction. Also, the users' subjective evaluations show significant improvement in many user-centric metrics. Experiences are discussed that shed light on avenues for future interface designs. <sup>1</sup>

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI); HCI design and evaluation methods**;

Additional Key Words and Phrases: Social Recommendation, User Interface, User-Driven Exploration, Diversification, Diversity-Promoting Interface

## ACM Reference format:

Chun-Hua Tsai and Peter Brusilovsky. 2010. Exploring Social Recommendations with Visual Diversity-Promoting Interfaces. *ACM Trans. Interact. Intell. Syst.* 9, 4, Article 39 (March 2010), 35 pages.

<https://doi.org/0000001.0000001>

## 1 INTRODUCTION

Recommending *people* within a social system is a challenging task. A traditional approach for offering recommendations is generating a single ranked list of relevant people that is adapted to the profile of the user requesting the information. The approach has been shown to be useful for solving the challenge of *social overload*, which is caused by the existing huge amount of digital data and social media content [24, 71]. A social recommender system aims to filter out “irrelevant” information so users can reduce the efforts of decision making. However, users may look for other

<sup>1</sup>This paper is an extended version of Chun-Hua Tsai and Peter Brusilovsky. 2018. Beyond the Ranked List: User-Driven Exploration and Diversification of Social Recommendation. In 23rd International Conference on Intelligent User Interfaces (IUI '18). ACM, New York, NY, USA, 239-250

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2009 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.

2160-6455/2010/3-ART39 \$15.00

<https://doi.org/0000001.0000001>

people for a range of reasons; for example, a conference attendee may wish to re-connect with an acquaintance, find new friends with similar research interests, or just find someone with whom to share a ride to the airport. This diversity of information needs makes it difficult to generate a static ranked social recommendation list that fits all cases [9].

A particular instance in which a single ranked list might not work well is in a parallel hybrid-recommendation system that fuses several recommendation sources. In this case, different sources might be preferred for various needs (i.e., social similarity could work best for finding known friends while content-based similarity could be used to find people with similar interests [9]). Several authors have argued that the best approach in this situation is to offer users the ability to control the fusion by choosing the contribution of various algorithms [9, 16] or data sources [5]. Providing a visual interface that makes the process of fusion more transparent - for example, by showing recommended sources and their overlaps as set diagrams [51, 67] - could further address this problem. However, the set-based approach that has previously been applied to visualize the results of controlled source fusion has limited applicability, since it ignores the strength of relevance within each of the sources (which is a continuous variable). In this paper, we attempt to overcome the limitations of set-based visual fusion by exploring two visual fusion approaches that reflect the continuously varying strength of relevance within each source while keeping the fusion process transparent.

When selecting a visual design for the transparent fusion of recommendation sources, we focused on better-informing users about the *diversity* of the recommended results. It has been previously demonstrated that a proper user interface could promote diversity in information exploration. For example, a diversity-enhancing interface evaluated in [29] led to higher user satisfaction than a simple ranked list interface. Several attempts to design a diversity-focused interface using a dimension reduction technique to present opinion similarity by latent distance have been presented in [17, 58, 70]. However, the clustering distance was not easily interpreted, and as a result, users were unable to make personalized judgments. It remains challenging to design a recommender interface in which the user can both perceive the diversity and be able to control the interface to filter the recommended items.

The project presented in this paper reports our experience with two different visual recommender interfaces. First, we proposed a recommender interface that explores the value of a two-dimensional scatter-plot visualization to present recommendations with several dimensions of relevance. In our context, the scatter plot interface was used to help users combine different aspects of relevance for recommended items, while providing inspectability to the users. Second, we proposed a recommender interface that enhances the fusion control function within a ranked list with meaningful visual encoding for multiple dimensions of relevance. The users can adjust the relevance weightings to customize the recommendation results, which provides the user with a greater level of control over their results.

The two interfaces were designed to explore the value of user-controllable and diversity-aware interfaces in a social recommender system. Each of the interfaces has been evaluated in a controlled field study in the target context. The results show that the new visual interfaces reduce exploration efforts for a set of realistic tasks, and also make the users more aware of the diversity of recommended items. Also, the users' subjective evaluation shows a significant improvement in many user-centric metrics. We further discussed the effects of the proposed interfaces on the users' experience with a diversity-enhanced social recommender system. A sequential analysis was performed to investigate the user interaction pattern while working with the two proposed visual recommender interfaces.

The paper offers several contributions: 1) We propose two interfaces that support the continuously controlled fusion of several relevance aspects with inspectability and controllability. 2) We provide

evidence that the diversity-aware interface not only helps the user to perceive diversity but also helps the user to improve usability in the real world beyond simple relevance tasks. 3) We discuss the user experience effects on proposed interfaces through a structural equation model analysis. 4) Finally, we examine and contrast user interaction patterns through a sequential mining approach. We report a set of frequent task-focused “micro patterns” for an in-depth analysis of user behavior.

## 2 BACKGROUND

### 2.1 Social Recommendation

Social media is a place to share various kinds of information as well as react to information shared, by using tags, likes, comments, or votes. The rapid growth of both content and data can be leveraged to provide accurate item recommendations, but it can also cause information overload, which makes it harder for users to filter interesting or relevant content [24]. This problem has been addressed in two different ways, forming two main streams or areas of research on *social recommendation*. The first stream focuses on improving the traditional recommender approaches (i.e., ranking-based collaborative filtering [35]) by using various kinds of social data available in social media systems such as social links [39], tags [3], or reviews [48]. The second stream focuses on people recommendations, which aims to reduce **social overload** on making people-to-people connections. Within social media, recommendations of “people you may know” [26] or new and interesting people to follow [25] could remarkably improve access to relevant information. However, the value of people recommendation is not limited to improving information access. For example, an online dating service can enable strangers to establish new personal relationships, while academic recommendation helps to find co-authors and project collaborators [55]. In this paper, we follow the second stream (people recommendation) and explore it in a primarily academic context.

People might need to connect for different reasons, and each reason might require different recommender approaches [25]. In this paper, we distinguish three common goals for people recommendation. First, **recommending people to connect with**: this kind of recommendation provides suggestions for users to establish *mutual* connections, which require an agreement from both parties. Mutual relationships vary in different systems; e.g., a relationship may be a friendship in a social networking service, professional networking in a job-matching website, or a connection in an event-based social system. For example, the event-based social system presented in [7], provided a “make a connection” function for connecting conference attendees and paper authors.

Since not all connections require mutual agreement, an important kind of recommendation is **recommending people to follow**. Its goal is to establish a follower-followee relationship between the users. For instance, on Twitter, a user can follow a friend, politician or movie star, which is a one-way relationship. The study of [23] attempted the use of a graph-based method to generate the “following” recommendations on Twitter, which applied random-walk techniques with a follower-followee network. Dynamic goals make it difficult to have a one-size-fits-all approach to all user needs. Moreover, relevance-driven goals may not be perceived or appreciated by the users [15]. The beyond-relevance objective is attracting more and more attention.

Finally, people recommendation is not limited to affirm or enforce existing social connections, but may aim to extend the social network to new connections, e.g., **recommending strangers** to the user. This is a useful function in many applications, such as seeking advice from an expert, finding new collaboration opportunities, or merely to explore new friends. [27] introduced a social networking service called *StrangerRS* that attempted to recommend an unknown user within a corporation. Their approach is to “present” the related information between two parties who may be potentially interested in connecting with each other. A user interface was designed and deployed

in a controlled field study. The experiment results showed the interface increases positive feedback among the recommended strangers.

## 2.2 Beyond-Relevance Objective

A conventional recommender system focuses on generating recommendations with high accuracy which is frequently driven by ranking or classification models. A model requires a training phase, based on collected user data, to recognize the high relevance items to the user. However, focusing solely on the high relevance items may not be enough to satisfy the various user needs [41]. The users may not even perceive a state-of-the-art recommender algorithm as useful. [15]. Moreover, an accuracy-driven RS may tend to deliver a narrow set of recommendation to the user, which is a well-known over-specification problem that leads to the poor user experience.

The factors of **beyond relevance** are attracting more and more attention in recent recommender system studies [2, 19, 30]. The literature recognizes four user-centered quality factors: 1) *Diversity*: generating recommendations that are “diverse” in different aspects. 2) *Serendipity*: providing useful recommendations, which go beyond user expectations. 3) *Novelty*: suggesting an unknown/new/long-tail item to a user, usually in contrast to the popular items. 4) *Coverage*: providing recommendations that cover multiple desired categories.

One way to achieve a better diversity of coverage is a hybrid recommender system that uses different recommender engines. However, using a hybrid recommender to generate a static ranked list doesn’t allow for the system to address changing user needs, which creates the challenge of increasing the system’s controllability, as well as making diversity in the recommendation items more evident [63]. At the same time, not every user values diversity equally [1]. The level of diversity-seeking is an existing individual difference. The findings of [43] indicated individual differences in the information-seeking process and the needs of designing a customized interface to fit different users. [6] argues for considering a “diverse conception of democracy” when we develop a diversity-enhancing tool or application. Furthermore, presenting only different information may not cause users to interact with diverse content. If a user feels threatened by unfamiliar information, a reinforcing effect may happen to cause users to avoid interacting with various types of content [17, 40]. Hence, a diversity-aware recommender system should consider item and user diversity, but not decrease the overall levels of user satisfaction and system usability [59].

## 2.3 Presenting Recommendation

There are studies tried to enhance the recommendation diversity through exploratory search interfaces. For instance, [49] introduced an interface for interactive search, which uses overlaying graphs representing different information sources. The studies [33, 34] proposed interfaces which can present multi-faceted information on map or touch screen to diversify and explore the generated recommendations. The studies of [70] adopted dimensional-reduction techniques to project the multidimensional data in two or three dimensions for the purposes of visualization. SciNet interface [20, 53] recommends keywords spatially in an interactive visualization to help diversification in exploratory search in a visualized surface of polar coordinates. In the field of information retrieval, a similar idea has been discussed to give users the capability in navigating and visualize the search results in multi-dimensional image renderer [13]. The work of [38] further adopted a 3D item space visualization for presenting multi-faceted data and user preferences in a movie recommender using collaborative filtering approach. The users can interactively manipulate the recommended item through adjusting the “landscape” on the map.

Some previous studies have been conducted to solve issues of recommendation diversity through interface design. For example, adopting a visual discovery interface can increase the click-through rate (CTR) across different item categories in an e-commerce website [58]. The user can explore

new or relevant products without the need for search queries. The key factor of the interface is to provide the user with the ability to control the filtering of the recommendation contents. [54, 70] proposed a user-controllable interface for users to interactively change the ranking or feature weighting for a better-personalized ranking. [68] proposed an interface to “tune” weighting of each tag genome. [17, 29] proposed interfaces to show the various recommendation results that promote the users’ perception of the diversity of recommendations. The study of [64] further demonstrated a more diverse exploration pattern when the user was interacting a two-dimensional interface for searching social connections.

Many scholars have suggested different explanation functions to increase the inspectability of recommender system [37]. The function provides the transparency that let users realize how the system works [28, 60]. The exposure of the recommendation process through visual interfaces can also increase the inspectability of the system [36]. Many different types of research have been done on this subject. For example, [65] provides recommendation visualization to increase the transparency of the recommender system. [67] provides a set-based visualization to let the user explore the desired recommendation items. Other researchers further indicated that the value of explaining interfaces could enhance user experiences. The explanation interface was associated with the perception of recommendation quality [60], gaining trust in the system [10] and experiencing the competence of the system [69]. The studies of [21, 46, 47] have mentioned that providing a controllable interface in the social recommender system can increase overall user satisfaction. The authors adopted interactive graphical interfaces to present the social recommendations that enable controllability of an item or user-level preference in a collaborative recommender system.

### 3 FIRST ATTEMPT: SCATTER VIZ

In a hybrid recommendation context with multiple types of relevance, the traditional ranked list makes it hard for the user to recognize how different relevance aspects are correlated. A typical example of this situation is recommending other attendees to meet at a research conference. Here a range of similarity functions (social, past publications, current interest, location) could indicate a person worth to meet. To help conference attendees in their conference networking, we proposed a dual social recommender interface, **Scatter Viz**, which includes a ranked list and visual scatter plot components. The ranked list was selected as a traditional way of presenting recommended results in a single dimension, listed from high to low relevance. The scatter plot was chosen as an intuitive way to present multidimensional data [32] and inspect patterns in large data-sets [12], as it has been shown that users can accurately judge data similarities between different shapes of scatter plots [50]. We hoped that the ability to view recommended items in two dimensions could reveal the overall diversity of results and help to correlate multiple types of relevance among social recommendations.

#### 3.1 Visual Design

Figure 1 illustrates the design of the dual interface in four sections.

- (1) **Section A** is the scatter plot. The interface presents each item (a conference attendee) as a circle on the canvas in two selected dimensions. The user can move the mouse over the circle to highlight the selection.
- (2) **Section B** shows the control panel, with which the user can interact. The user can select the number of recommendations to display and choose the *major feature* and the *extra feature* to visualize the recommendations on the scatter plot. The major feature is used to rank the results along the X-axis of the Scatter Plot (Section A) and in the ranked list (section C).



Fig. 1. The design of Scatter Viz: (A) Scatter Plot; (B) Control Panel; (C) Ranked List; (D) User Profile Page. The interface supports exploration of recommended items in Section A or C and detail inspection in section D. The scholar names have been pixelated for privacy protection.

The extra feature is used to diversify the recommendations in respect to the selected aspect along the Y-axis (thus spreading the results that have similar values of the major feature but different values of the extra feature). To further investigate the diversity of the displayed recommendations, the user can also use another data aspect as a *category* to color-code the results. The default category was *Smart Balance*, which highlights four quadrants of the displayed data with a 0.5 ratio. Figure 2 presents six sample scatter plot layout combining “Academic”, “Social”, “Interest” and “Distance” relevance features that are color-coded using the Smart Balance Category.

- (3) **Section C** is the standard ranked list. More precisely, it is a combination of four ranked lists produced by four recommender engines, as explained below. To make the four dimensions more comprehensive, the model normalized relevance scores from 0-1 of each user to the target user, generated by each recommender engine. All the relevance scores are shown on the right side of the ranked list. The user can hover over each row to highlight the location in the scatter plot or click for a more detailed user profile.
- (4) **Section D** presents more detailed information about the person who has been selected in either the visualization or the ranked list. Among other aspects, four of the six tabs visually explain how each recommender engine calculates the relevance of the selected user to the target user. Due to the page limitation, the details of each explanation tabs are omitted. The design detail of the explanation functions can be found in [65].

The visual encoding affects the way users process the information. Pre-attentive processing let users absorb and precept the enormous amount of information in a short period [14]. The proposed interface helps to present the recommendation results in two kinds of visual encoding. First, the interface displays the recommendation relevance in two dimensions. The visual encoding helps the user to *spot* the item in different dimensions. It helps the user to make a decision beyond single or

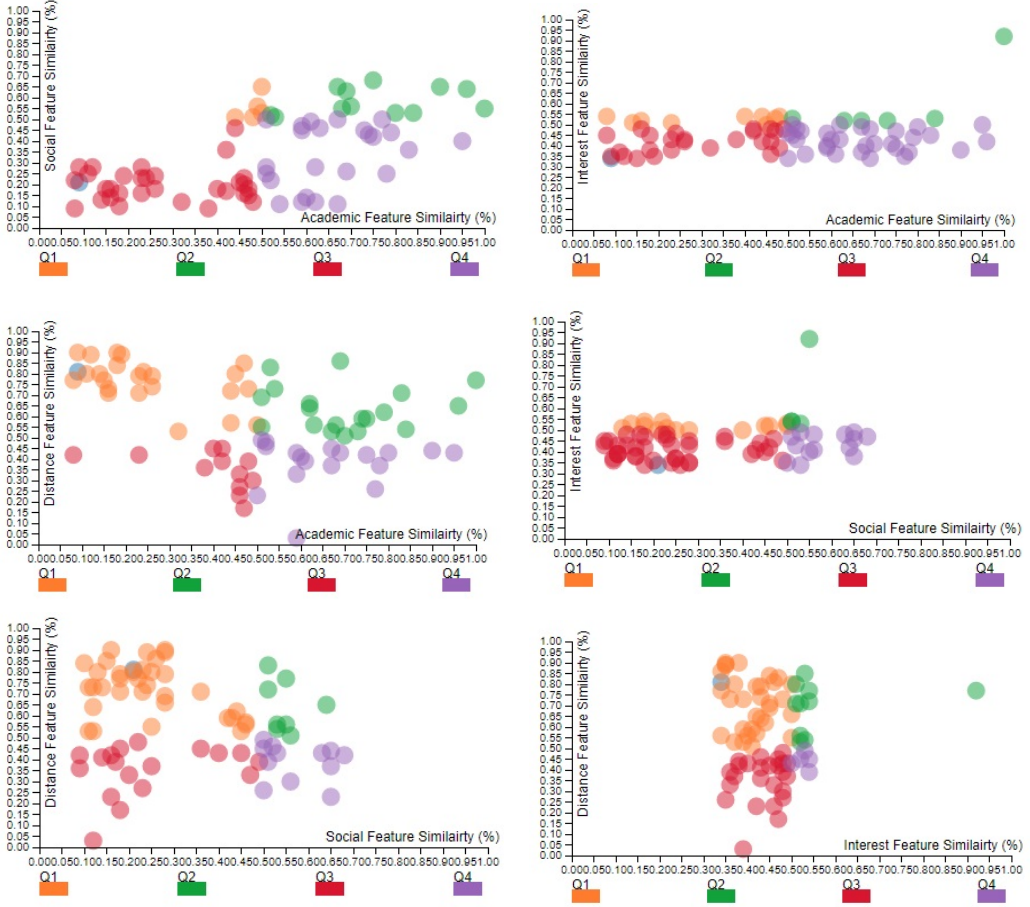


Fig. 2. Scatter plot layouts: the layout would adjust, based on the selected *Major* and *Extra* features (Section B in Fig. 1). Here is an example that presents the same recommended items in six dimension combinations: Academic/Social, Academic/Interest, Academic/Distance, Social/Interest, Social/Distance and Interest/Distance feature coordinates. The nodes are colored using four equal quadrants that set Category as “Smart Balance”.

combined relevance, which is more realistic in many real-world scenarios. For example, a user may be interested in a scholar whose research area is highly relevant to their research and who is also affiliated with nearby cities. The scatter interface helps to filter a group of recommended items with the two desired relevance features. Second, the node is color-coded in different categorical features; for example, in Smart Balance mode, the node is color-coded by the four quadrants between two dimensions of features. The user can perceive the tendencies of the recommendation item, based on their coloring, and the user can also update the layout with different Category features, including the meta-data of the recommended scholar’s title, position, and home country. In addition, both the node and table row are highlighted synchronously while the user moves over the recommended items (see example in Figure 1). As a result, the scatter plot interface can be used for recommended item selection or just as a diversity-oriented recommendation explanation.

### 3.2 Personalized Relevance Model

To rank the recommended attendees by their relevance to the target user, the system uses four separate recommender engines that rank other attendees along four dimensions that we call *features*: text similarity of their *academic* publications, *social* similarity through the co-authorship network, similarity of current *interests* measured as similarity of their bookmarked talks, and the *distance* of their place of affiliation to the target user. Each of these features is defined below.

- **The Academic feature** is determined by the degree of publication similarity between two attendees using cosine similarity [66]. The function is defined as:

$$Sim_{Academic}(x, y) = (v_x \cdot v_y) / \|v_x\| \|v_y\| \quad (1)$$

where  $v$  is term vectors for user  $x$  and  $y$ . We used *tf-idf* to create the vector with a term frequency (*tf*) upper bound of 0.5 and lower bound of 0.01 to eliminate both common and rarely used terms. The *tf-idf* function is defined as below.

$$tf-idf(t, d, D) = tf(t, d) \times idf(t, D) \quad (2)$$

$$idf(t, D) = \ln\left(\frac{N}{|\{d \in D : t \in d\}|}\right) \quad (3)$$

where  $t$  is the given term in document  $d$ ,  $D$  represents all documents in the corpus,  $|\{d \in D : t \in d\}|$  is the number of documents where the term  $t$  appears and  $N$  is total number of documents in the corpus.

- **The Social feature** approximates the social similarity between the target and recommended users by combining co-authorship network distance and common neighbor similarity from publication data. We adopted the depth-first search (DFS) method to calculate the shortest path  $p$  [57] and common neighborhood (CN) [44] for the number  $n$  of coauthors overlapping in two degrees for user  $x$  and  $y$ . A longer path (larger  $p$ ) is presenting less similar co-authorship network, so we used a constant  $C = 20$  to reverse the weighting.

$$Sim_{Social}(x, y) = (C - p) + n \quad (4)$$

- **The Interest feature** is determined by the number of co-bookmarked papers and co-connected authors within the experimental social system [7]. The function is defined as:

$$Sim_{Interest}(x, y) = (b_x \cap b_y) + (c_x \cap c_y) \quad (5)$$

where  $b_x, b_y$  represent the bookmarked papers of user  $x$  and  $y$ ;  $c_x, c_y$  represents the friend connections of user  $x$  and  $y$ .

- **The Distance feature** is a measure of geographic distance between attendees. Based on attendees' affiliation information, we retrieve longitude and latitude coordinates from Google Map API. We used the Haversine formula to compute the geographic distance between any pair of attendees [66]. The formula is defined as:

$$Sim_{Distance}(x, y) = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_y - \varphi_x}{2}\right) + \cos(\varphi_x)\cos(\varphi_y)\sin^2\left(\frac{\lambda_y - \lambda_x}{2}\right)}\right) \quad (6)$$

where  $r$  is the radius of the sphere;  $\varphi_x$  and  $\varphi_y$  are latitude of user  $x$  and  $y$ 's affiliation;  $\lambda_x$  and  $\lambda_y$  are longitude of user  $x$  and  $y$ 's affiliation.

### 3.3 Diversity Navigation Model

The Personalized Relevance Model determines the combined relevance score for each conference attendee, i.e., instead of ranking the recommended people using a static ensemble fusion of relevance aspects, the system allows the user to rank and visualize items using on different aspects of relevance through our proposed interface. We can measure the user's selection diversity with two diversification models.

- (1) **Feature Diversification:** the user can select any two pairs of proposed features and spot the recommended items from the intersection of their relevance. All of the proposed features were calculated on a different scale. For example, the distance feature is the physical distance in miles, while the academic feature is calculated as a percentage. To enable the comparison of diverse features, we adopted a standard Z-score to normalize all the features to the same scale, from 0 to 1. The function was defined as:

$$ZScore = \frac{x_i - u_j}{\sigma_j} \quad (7)$$

where  $x_i$  is the  $i$ th recommended item and  $j$  represents the corresponding feature with its average  $u$  and variance  $\sigma$ . Then, we use a standard Z-table to convert the  $ZScore$  to the corresponding percentile  $p_{ij}$ . Hence, we can list all the features on the same scale for presentation in a ranked list and scatter plot diagram.

- (2) **Category Diversification:** it is a model of diversifying the different categories [31]. In the scatter plot, we color-code the items from different categories, such as title, position, and country. In the ranked list, we listed the category as one column for a user to access.
- (3) **Selection Diversity:** we can then measure the user selection/exploration diversity, based on the two diversification models. We observe the user's interaction with items from different "quadrants" (feature intersections) [61], such as high academic and high social features, or high academic and low social features. The extent of diversity is measured by *Shannon Entropy*:

$$Entropy : d_u = - \sum_{i=1}^4 p_i \log_4 p_i \quad (8)$$

where  $p_i$  is the probability for a particular quadrant (feature or category) and the proportion of all of the user's selections [42]. Based on the definition, we can measure the diversity in the different aspects of the relevance dimension. We can compare the combinations of all the proposed features. For example, in a recommendation system fused with *four* features. We can measure the entropy difference among the  $4 * (4 - 1) = 12$  pair of dimensions.

## 4 STUDY 1: RANKED LIST V.S. SCATTER VIZ

### 4.1 Data and Participants

The recommendations produced by all four engines are mostly based on data collected by the Conference Navigator 3 (CN3) system [7]. The system has been used to support 38 conferences at the time of writing this paper and has data on approximately 6,398 articles presented at these conferences; 11,939 authors; 6,500 users (attendees of these conferences); 28,590 bookmarks; and 1,336 social connections. We used the Aminer dataset [56] to mediate the cold-start issue for academic and social engines that occurs when users have no publications or co-authorship information [62].

A total of 25 participants (13 female) were recruited for the user study. All of the participants were attendees at the 2017 Intelligent User Interfaces Conference (IUI 2017). Since the primary goal of our system was to help junior scholars connect with other people in the field, we specifically selected junior scholars, such as graduate students or research assistants. The participants came from 15 different countries; their ages ranged from 20 to 50 ( $M=37$ ,  $SE=7.07$ ). All of them could be considered as knowledgeable in the area of the intelligent interface for at least one academic publication from IUI 2017. To control for any prior experience with the recommender system, we included a question about in the background questionnaire. The average answer score was ( $M=3.28$ ,  $SE=1.13$ ) on a five-point scale.

### 4.2 Experiment Design and Procedure

To assess the value of the proposed interface, we compared the dual Scatter Viz interface with the scatter plot and the ranked list (we will call this condition as SCATTER) with a baseline interface using only a ranked list (RANK) with Section A (in Figure 1) removed. The study used a within-subjects design. All participants were asked to use each interface for three following tasks and to fill out a post-stage questionnaire at the end of their work with each interface. At the end of the study, participants were asked to compare interfaces regarding their explicit preference. The order of using interfaces was randomized to control for the effect of ordering. In other words, half of the participants started the study with the SCATTER interface. To minimize the learning effect (becoming familiar with data), we used data from two years of the same conference: the SCATTER interface used papers and attendees from IUI 2017, while the RANK interface used the corresponding data from IUI 2016.

Participants were given the same three tasks for each interface. The tasks were explicitly designed as diverse but realistic tasks that could be naturally pursued by attendees at research conferences.

- **Task 1:** Your Ph.D. adviser has asked you to find four Committee Member candidates for your dissertation defense. You need to find candidates with expertise close to your research field while trying to lower their travel cost to your defense.
- **Task 2:** Your adviser has asked you to meet four attending scholars, preferably from different regions across the world, who have a close connection to your research group.
- **Task 3:** You want to find four junior scholars (not yet faculty members) with reasonably similar interests among the conference attendees to establish networking.

The participants were asked to pick suitable candidates among conference attendees, based on their best judgment in each task. When designing the tasks, we attempted to make them realistic, yet focused on multiple aspects of relevance, as many real tasks are. We consider that task 1 is relevance-oriented and that tasks 2 & 3 are diversity-oriented. For a relevance-oriented task, we expect to see if the proposed interface helps the user to coordinate different relevance aspects of the desired target efficiently. In contrast, for the diversity-oriented task, we expect the system to help to recognize the diversity of recommended items, as compared to the baseline interface.

| Task | Action          | RANK            |            | SCATTER         |            |
|------|-----------------|-----------------|------------|-----------------|------------|
|      |                 | M (SE)          | User Count | M (SE)          | User Count |
| T1   | Control Panel   | 3.88 (2.40)     | 24         | 4.12 (2.02)     | 25         |
|      | Explanation Tab | 34.28 (29.50)   | 25         | 7.96 (7.48)     | 19         |
|      | Click - Rank    | 26.28 (29.50)   | 25         | 4.92 (6.75)     | 15 *       |
|      | Click - Scatter | -               | -          | 3.04 (5.45)     | 13 *       |
|      | Time Spending   | 345.44 (209.86) | 25         | 389.12 (235.29) | 25         |
| T2   | Control Panel   | 2.88 (1.64)     | 24         | 2.88 (1.12)     | 25         |
|      | Explanation Tab | 19.96 (17.47)   | 25         | 9.16 (6.28)     | 25         |
|      | Click - Rank    | 16.96 (17.47)   | 25         | 2.68 (4.69)     | 15 *       |
|      | Click - Scatter | -               | -          | 3.48 (5.41)     | 13 *       |
|      | Time Spending   | 216.6 (144.95)  | 25         | 190.84 (115.33) | 25         |
| T3   | Control Panel   | 2.56 (1.04)     | 24         | 2.84 (1.10)     | 25         |
|      | Explanation Tab | 20.08 (20.29)   | 25         | 6.4 (7.22)      | 19         |
|      | Click - Rank    | 19.08 (20.29)   | 25         | 3.48(7.80)      | 9 *        |
|      | Click - Scatter | -               | -          | 2.92 (2.95)     | 15 *       |
|      | Time Spending   | 345.95 (156.39) | 25         | 369.2 (169.77)  | 25         |

Table 1. User action summary of study 1: the table shows the user interaction statistics while performing each of the three tasks using two interfaces. (Statistical significance level: (\*)  $p < 0.05$ .)

### 4.3 Action Analysis

Table 1 shows the system usage for two interfaces. The data indicate that participants extensively used both the control panel and explanation tabs to complete the tasks. The participants usually required more actions on the first task to familiarize themselves with the system. There is no significant difference on the action of change control panel and the click on the explanation tab between the interfaces in three tasks; although, in the SCATTER interface, the users tended to click the explanation functions less. The click frequency presented a significant difference between the two interfaces. This finding is not surprising because the RANK interface lacks the visualization information that pushes the participant to click more on the user profile page to inspect the necessary information. It is interesting to see that not every user clicked on the scatter plot. This data hints that some participants treated the scatter plot visualization as an explanation function rather than an interactive exploration interface. At the same time, we found no significant difference in the time spent on the tasks. The data hints that each action taken in the SCATTER interface delivered more interesting information with which to engage.

It would be valuable to see how the participants adopt the control and explanation functions. Figure 3 presents the usage of four similarity features and three category features. The red line shows how frequently the feature was selected as the primary ordering factor. The factor determines 1) the ranking in the RANK interface, and 2) the x-axis layout in the SCATTER interface. The purple line shows the frequency of its use as the second feature to spread the results along the y-axis in the SCATTER interface. The blue line is the frequency of feature used as the Category feature. The chosen category feature creates one additional column in the RANK interface and updates the color-coding in the SCATTER interface. We can observe the usage pattern between the three proposed tasks. In task 1, which focuses on finding the 2-category optimum ranking, the participants most frequently used Academic as the primary ranking feature, diversifying it by Distance and selecting the Position feature to color-code the results. In diversity-oriented task 2, a number of diversification features were tried with about the same frequency. In task 3 (find

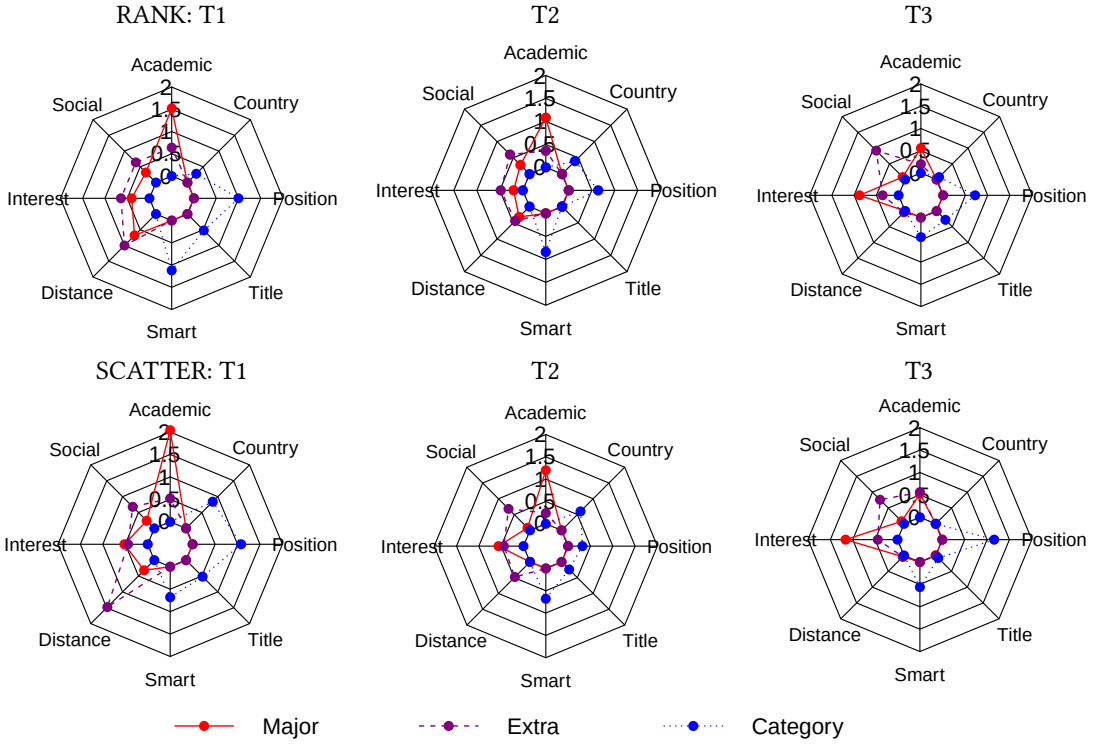


Fig. 3. The recommendation features usage of Study 1. The red line is the usage of major features, the purple line is the usage of extra features, and the blue line is the usage of category features.

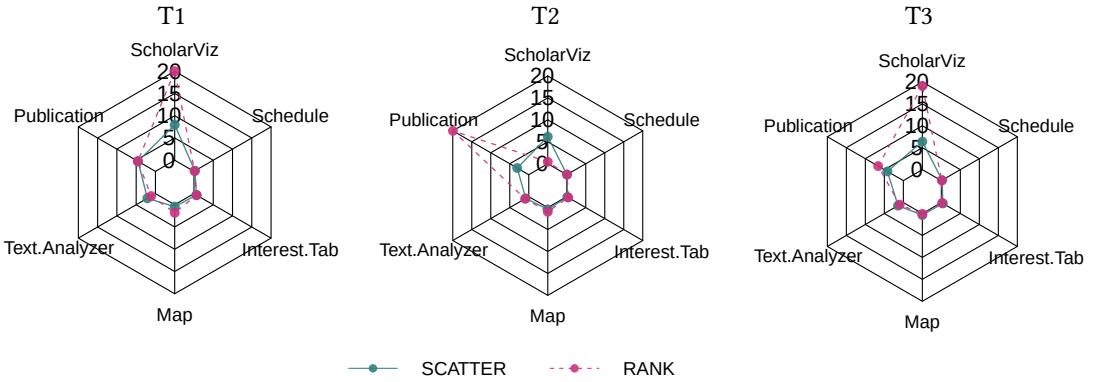


Fig. 4. The explanation function usage of Study 1. The light plum color is the usage of the RANK interface, while the light blue color is the usage of the SCATTER interface.

four junior scholars with similar interests), the Interest feature was most frequently selected as the primary feature, while Social was the primary diversification approach and Position was the primary color-coding approach. The pattern was consistent between two interfaces. Overall, this result shows that the users were quite efficient in selecting the most useful features for each

task. Figure 4 shows the click frequency on six different explanation tabs. Overall, the users were most interested with explanations presented in ScholarViz and Publication tabs (which explained the Social and Academic similarities, respectively). The figure also shows that explanations were requested more frequently for candidates accessed using the RANK interface.

#### 4.4 User Feedback Analysis

To compare subjective feedback, we analyzed the responses of the post-stage questions using paired sample t-tests. Figure 5 shows the result of this analysis. We compared the twelve aspects of subjective feedback from the participants; among them, the SCATTER interface received a significantly higher rating for six aspects: Trust (Q4), Supportiveness (Q5), Interest (Q6), Satisfaction (Q8), Intention to Reuse (Q9), and Enjoyable (Q11). In two questions, facilitation (Q7) and the Reversed Benefit Question (Q12), the SCATTER interface scored higher, but not significantly so. It is interesting to see that the RANK interface scored a bit higher (though not significantly so) on explanation usefulness, which hints that the lack of visualization made explanations more important in the RANK interface. In the final preference test, the SCATTER interface received much stronger support than the RANK interface in the user preference feedback (Figure 6). Most importantly, a majority of users (84%) considered the SCATTER interface to be a better system for recommending attendees and better help in diversity-oriented tasks, as well as a better system for recommending.

#### 4.5 Selection Diversity Analysis

Table 2 shows the diversity analysis for each task and interface. The result shows the users' responses to the tasks with a different pattern of exploration, which caused a variance of diversity and coverage measurements. All three tasks are shown a least one significance between two interfaces but in the different aspects of features. For the SCATTER interface, task 1 (relevance-oriented) shows significance statically on less difference between academic/social & social/interest features, but more coverage on the title category. Tasks 2 & 3 (diversity-oriented) show higher selection diversity in the interest/distance and social/distance features, respectively, as well as higher selection coverage in the title & country category features. The data supports the finding that the SCATTER interface helped the participants to accurately filter the attendees in the relevance-oriented task, as well as extend the selection diversity in the diversity-oriented tasks.

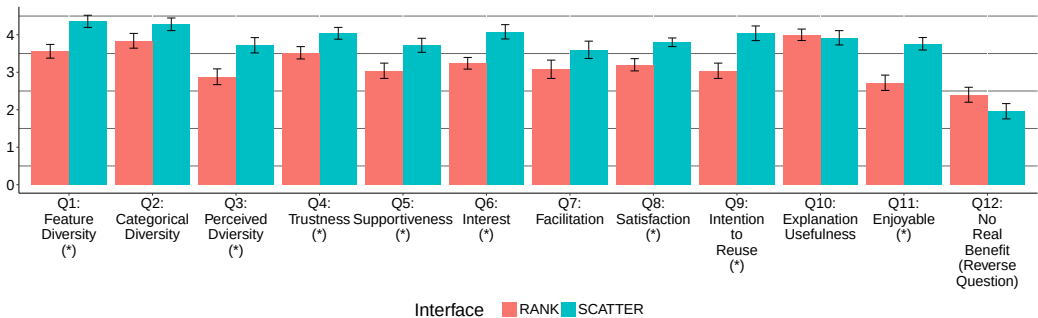


Fig. 5. User feedback analysis results for Study 1 shows that the SCATTER interface received a significantly higher rating for six aspects. (A cut-off value was set at 3.5 on the 5 point scale. Statistical significance level: (\*)  $p < 0.05$ .)

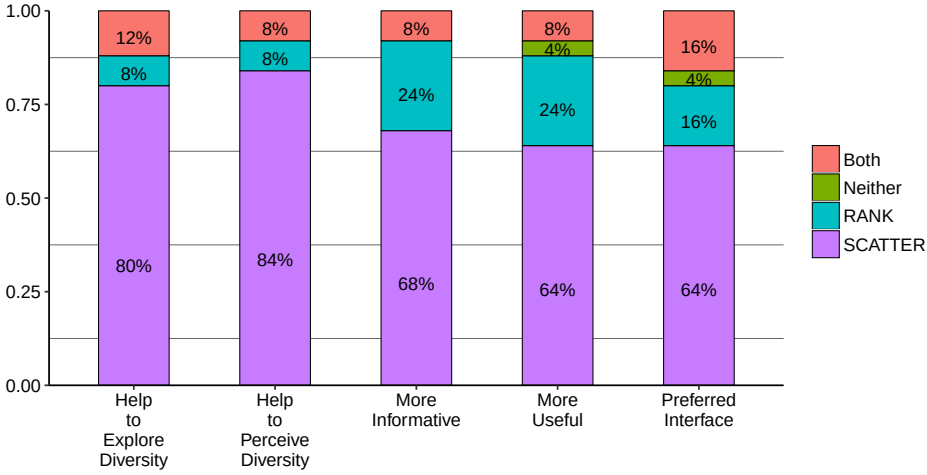


Fig. 6. User preference analysis in study 1 (the preferences were collected after the users experienced both interfaces). The result shows the SCATTER interface was preferred by the users in all aspects.

| Dimensions          | Task 1         |                   |   | Task 2         |                   |   | Task 3         |                   |   |
|---------------------|----------------|-------------------|---|----------------|-------------------|---|----------------|-------------------|---|
|                     | RANK<br>M (SE) | SCATTER<br>M (SE) | P | RANK<br>M (SE) | SCATTER<br>M (SE) | P | RANK<br>M (SE) | SCATTER<br>M (SE) | P |
| Academic + Social   | 0.14 (0.06)    | 0.11 (0.03)       | * | 0.14 (0.06)    | 0.13 (0.04)       |   | 0.14 (0.06)    | 0.13 (0.03)       |   |
| Academic + Interest | 0.16 (0.09)    | 0.13 (0.08)       |   | 0.14 (0.09)    | 0.14 (0.08)       |   | 0.12 (0.07)    | 0.12 (0.08)       |   |
| Academic + Distance | 0.13 (0.07)    | 0.12 (0.04)       |   | 0.12 (0.07)    | 0.14 (0.04)       |   | 0.14 (0.07)    | 0.16 (0.04)       |   |
| Social + Interest   | 0.27 (0.13)    | 0.21 (0.09)       | * | 0.25 (0.13)    | 0.24 (0.11)       |   | 0.22 (0.12)    | 0.22 (0.14)       |   |
| Social + Distance   | 0.27 (0.12)    | 0.24 (0.08)       |   | 0.26 (0.13)    | 0.28 (0.10)       |   | 0.24 (0.11)    | 0.31 (0.10)       | * |
| Interest + Distance | 0.26 (0.13)    | 0.25 (0.13)       |   | 0.23 (0.14)    | 0.27 (0.13)       | - | 0.21 (0.14)    | 0.23 (0.11)       |   |
| Title               | 0.17 (0.12)    | 0.22 (0.07)       | * | 0.17 (0.14)    | 0.31 (0.12)       | * | 0.17 (0.16)    | 0.32 (0.10)       | * |
| Position            | 0.26 (0.12)    | 0.23 (0.10)       |   | 0.29 (0.17)    | 0.25 (0.14)       |   | 0.19 (0.14)    | 0.15 (0.08)       |   |
| Country             | 0.49 (0.25)    | 0.46 (0.15)       |   | 0.46 (0.31)    | 0.68 (0.26)       | * | 0.44 (0.26)    | 0.66 (0.26)       | * |

Table 2. Diversity analysis for study 1: the table shows selection diversity for three tasks in the feature and category dimensions. The result shows that the SCATTER interface can help users to explore a more diverse set of recommendation in diversity-oriented tasks (T2 & T3). (Statistical significance level: (\*)  $p < 0.05$ ; (-)  $p < 0.1$ .)

#### 4.6 Discussion

In study 1, we evaluated a dual visual interface for recommending attendees at a research conference. A research conference context introduces several dimensions of attendee relevance, such as social, academic, interest, and distance similarities. Due to these factors, a traditional ensemble ranked list makes it difficult to express the diversity of recommended items (attendees). By spreading rankings over two dimensions, the suggested interface helps users to explore recommendations and recognize their diversity in several aspects. To assess the visual approach, we conducted a user study in a real conference environment to compare our interface (SCATTER) with a traditional ranked list (RANK) in three practical tasks. Our experimental result shows a tangible incremental impact on the metrics of system usage, efficiency, and diversity.

We found that the **Scatter Viz** interface can improve user inspection on recommendations with multi-relevance, which leads to a higher selection diversity in the given tasks. However, we also noticed that some of the experiment participants still stick to the familiar ranked list, even when an enhanced visualization was provided. This finding helps us to realize a user preference on adopting the interface with lower learning efforts. Besides, the scatter visualization requires additional space to present, which may not be feasible in many real-world rank-based recommender systems. These findings lead to our second attempt at extending the ranked list with multi-aspect awareness, controllability, and diversity-aware designs.

## 5 SECOND ATTEMPT: RELEVANCE TUNER

The ranked list is widely applied to recommender systems for presenting recommendations to users. Even in visual recommender systems, a basic ranked list is still essential for user interactions [51, 63, 67]. A recommender system usually ranks recommended items from high to low relevance, which may reduce users' cognitive decision loading [4]. However, as mentioned above, in a context with multiple relevance aspects, a statically "ensembled" ranked list makes it more difficult for the user to recognize the impact of different aspects and to adjust the recommendations to different needs. Moreover, the persuasive design of the ranked list causes users to pay more attention to items on top of the list [11], which further decrease selection diversity and caused the fitter bubble effect though narrowing the recommendation selection [45].

As the first study shows, the users were able to properly combine relevance dimensions in a two-dimensional scatter plot visualization, but that this ability came with a steep learning curve. Despite benefits offered by the Scatter Plot, the more familiar Ranked List component was used more heavily. In our second attempt, we explored a novel interface that allowed users to explore multiple aspects of relevance within the familiar ranked list extended with a controllable tuner and stackable color bars. We proposed the **Relevance Tuner** - a visual interface with user-driven control function and meaningful visual encoding. This design expanded the ranked list representation with an ability to visualize and control multiple aspects of item relevance, which is especially important for diversity-oriented tasks. The rank-based design reduces the user's learning efforts in getting familiar with the interface.

The design is inspired by several previous research works. Ekstrand et al. [16] argued for the need to use multiple recommendation algorithms within a single system. Their approach is to let users choose the algorithm based on specific information needs. For instance, the content-based method performs better on exploring new friend while the collaborative-filtering approach is out-performed others on re-connecting old friends [9]. A social recommender system that supports multiple information needs should be controllable. In many real-world scenarios, the user might need to *fuse* multiple methods or data sources to fulfill their information needs. For example, a controllable slider has been adopted for fusing and filtering different recommendation features, i.e., information sources, algorithms, user tags and keywords [5, 14, 51, 68]. Various kinds of stackable bars were suggested to visualize sources of algorithms that used to support a specific recommendation [8, 14, 51]. When designing **Relevance Tuner**, we intend to support source fusion with a set of relevance sliders and proportional stackable bars. The design of the Relevance Tuner is shown in Figure 7.

- (1) **Section A** contains five controllable sliders with the different colors representing the features of the Personalized Relevance Model. The scale of the slider ranges from 0 to 10. The user can change the weighting ( $W$ ) on the fly to re-rank the ranked list below. It provides controllability for the user to adjust the ranking to different recommendation needs and preferences. The interface also adds one new feature: Social Context. This feature computes

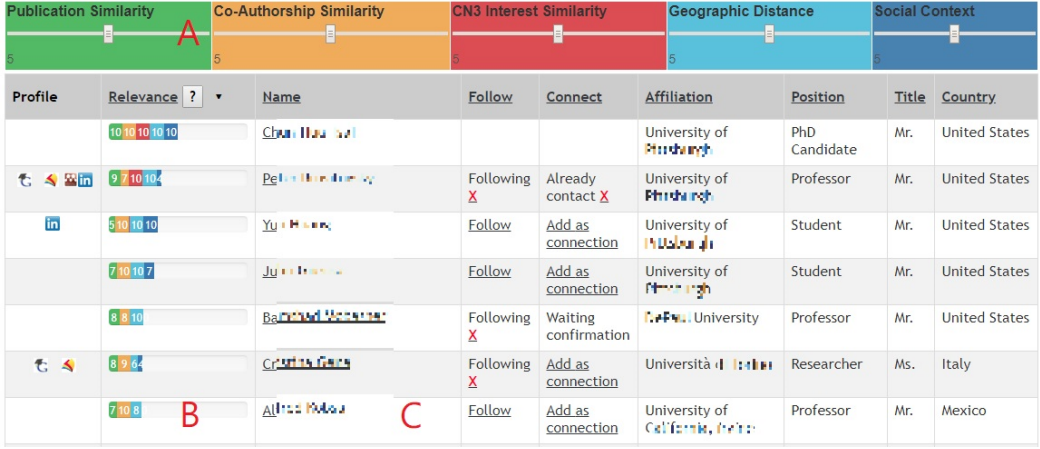


Fig. 7. The design of the Relevance Tuner: (a) Relevance Sliders; (B) Stackable Score Bar; (C) User Profiles. The interface enables the user to adjust the feature weighting on-the-fly for retrieving a customized recommendation list. The user can examine the relevant aspects of the recommended item through the multicolored score bar. (The Name and Affiliation entities have been pixelated for privacy protection.)

the Google search result based on the scholar's name and affiliation information; that is, the text similarity of the homepage and other related search results.

- (2) **Section B** shows the stackable relevance score bar for each recommended item in the ranked list. A stackable color bar interface is known for its ability to enhance controllability and transparency in a multi-aspect ranking [14]. In our system, the stackable color bars help the user to perceive how different relevance aspects of a recommended item are combined while adding transparency to the multi-aspect recommendation process. Each colored segment in the bar corresponds to one of the relevance aspects indicated by its color, which corresponds to one the sliders in **section A**. The length of the segment can vary between 0 and 20 and is determined by both, item relevance within this aspect (feature similarity) and the importance of this aspect (weight) selected by the user by adjusting sliders. The relevance score ( $R$ ) is defined as:

$$R_{ij} = \text{Round}\left(\left(\frac{f_{ij}}{\max_j} * 10\right) * \frac{W_j}{5}\right) \quad (9)$$

where  $f_{ij}$  is the  $i$ th recommended item's feature similarity for the aspect  $j$ ,  $\max_j$  is the local maximum value of  $j$ , and  $W_j$  represents the current slider weight for  $j$ . We use the weighting percentage ( $\frac{W_j}{5}$ , ranged from 0 to 2) to convert the normalized feature similarity score ( $\frac{f_{ij}}{\max_j} * 10$ , ranged from 0 to 10) to the corresponding relevance score ( $R_{ij}$ , ranged from 0 to 20). For example, in Figure 7, the top-ranked recommendation's *Normalized Publication Similarity Score* is 10. If a user changes the *Publication Similarity* slider weight ( $W$ ) to 8, then the relevance score would be  $10 * \frac{8}{5} = 16$ . The second recommendation's relevance score would be  $9 * \frac{8}{5} = 14.4$ . We will get the final relevance score as 15 after the roundup function. All relevance scores in each row will be updated, the entire list will be re-ranked by the sum of five relevance scores.

- (3) **Section C** shows the recommended scholar's meta-data, including name, social connection, affiliation, position, title, and country. The user can sort the ranked list by clicking the head of each column, or can inspect the explanation tabs (same as Section C in Figure 1) by clicking the name entities.

## 6 STUDY 2: SCATTER VIZ V.S. RELEVANCE TUNER

### 6.1 Data and Participants

Study 2 was conducted through the Conference Navigator 3 (CN3) system. The data was extended from Study 1 to a new conference: the 25th Conference on User Modeling, Adaptation, and Personalization (UMAP 2017). A total of 20 participants (7 female) were recruited for the user study. All of the participants were attendees at the UMAP 2017 conference. They were from 15 different countries; their ages ranged from 20 to 40 ( $M=31.19$ ,  $SE=4.97$ ). All of them had at least one publication from UMAP 2017. The background knowledge of recommender systems score was ( $M=3.85$ ,  $SE=0.79$ ) on a five-point scale.

### 6.2 Experiment Design and Procedure

In Study 2, we compared the interface of the ranked list plus the relevance tuner (TUNER) with a baseline of the scatter plot plus ranked list (SCATTER). The experiment design and procedure repeat the setting of Study 1. We manipulated the new proposed interface and adapted data from different conferences: the SCATTER interface used papers and attendees from UMAP 2017, while the TUNER interface used the same data from UMAP 2016, to minimize the learning effect between the two manipulations.

### 6.3 Action Analysis

Table 3 shows the system usage for the two interfaces of Study 2. In TUNER interface, the control panel usage is defined as each time the user moves the sliders. The data supports the users interacting more frequently (it shows significance in all three tasks) with the control panel in TUNER than in SCATTER. Conversely, the users clicked more on explanation tabs in SCATTER than in TUNER. The data implies that the information listed on the table was sufficient for the users to inspect and make decisions in three proposed tasks. In task 1, the SCATTER has a significantly higher clicking frequency and longer time spent (not significant) than the TUNER interface. The same pattern repeats in task 2 & 3, which shows that the users took more time to get familiar with the SCATTER interface. The users were gaining familiar with the TUNER interface more rapidly than with the SCATTER interface.

The analysis of control and explanation function usage is reported in Figures 8 & 9 and 10. Figure 8 shows the re-weighting frequency of the TUNER interface. The red line is the usage of Tuner sliders, and the blue line is the average feature score selected by the user during re-tuning. The data indicates that the user tends to interact with the sliders more in the beginning (task1), but interacts less in a later task (2 & 3). However, the overall amount of manipulation with features was remarkably higher for TUNER than SCATTER (Figure 9) for all tasks. The average weighting scores for TUNER show only a very slight association with the nature of the task at hand; that is, all features were potentially useful to filter the recommendation result in both relevance and diversity-oriented tasks. In contrast, the relatively rare use of SCATTER features (Figure 9) showed a consistent pattern with Study 1 - the users selected the recommendation features based on the tasks' requirement. The comparative use of explanations shown in Figure 10 is also consistent with Study 1 showing that the main demand was for *socialviz* and *publication* tabs. At the same time, it is interesting to observe that users in the TUNER group requested explanation less frequently

| Task           | Action          | TUNER        |            | SCATTER       |            |
|----------------|-----------------|--------------|------------|---------------|------------|
|                |                 | M (SE)       | User Count | M (SE)        | User Count |
| T <sub>1</sub> | Control Panel   | 38.4 (37.71) | 20         | 2.85 (2.23)   | 18         |
|                | Explanation Tab | 9.35 (8.28)  | 20         | 22.95 (23.71) | 20 *       |
|                | Click - Rank    | 5.05 (2.45)  | 20         | 9.8 (8.43)    | 17 *       |
|                | Click - Scatter | -            | -          | 4.1 (6.03)    | 11         |
|                | Time Spending   | 357 (289.04) | 20         | 537 (596.98)  | 20         |
| T <sub>2</sub> | Control Panel   | 15.2 (13.63) | 19         | 2 (1.71)      | 17 *       |
|                | Explanation Tab | 6.5 (8.74)   | 20         | 8.45 (6.79)   | 20         |
|                | Click - Rank    | 4.3 (1.21)   | 20         | 5.2 (3.76)    | 18         |
|                | Click - Scatter | -            | -          | 1.8 (2.94)    | 8 *        |
|                | Time Spending   | 201 (235.43) | 20         | 294 (470.78)  | 20         |
| T <sub>3</sub> | Control Panel   | 12.2 (11.67) | 17         | 2.25 (1.80)   | 19 *       |
|                | Explanation Tab | 9.25 (8.75)  | 20         | 12.9 (14.38)  | 20         |
|                | Click - Rank    | 5.15 (2.51)  | 20         | 10.2 (16.93)  | 17         |
|                | Click - Scatter | -            | -          | 2.05 (3.13)   | 10 *       |
|                | Time Spending   | 153 (92.28)  | 20         | 285 (470.78)  | 20         |

Table 3. User action summary of study 2: the table shows the statistics of user interaction while solving each of the three tasks using each interface. (Statistical significance level: (\*)  $p < 0.05$ .)

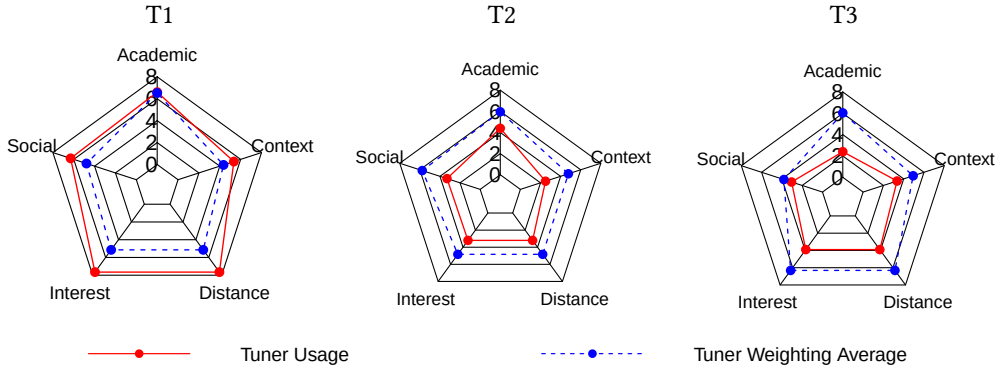


Fig. 8. The relevance slider usage of Study 2. The red line is the tuner usage, while the blue line is the average weighting score of the overall re-tuning.

than the SCATTER group, as through a regular TUNER interface provided more information for decision making than the SCATTER interface. However, we also observed that the TUNER users took less time to finish the tasks, as compared to those using the SCATTER interface.

#### 6.4 User Feedback Analysis

Figure 11 shows the analysis of the post-stage survey. The high rating in both interfaces shows the positive user acceptance in Study 2 (no significance on all the factors). However, the user tends to favor the TUNER interface when considering the factors of Supportiveness (Q5), Interest (Q6), Facilitation (Q7), Satisfaction (Q8), Intend to Reuse (Q9), and Usefulness (Q10). The SCATTER interface performs better on the measures of Trustiness (Q4) and Enjoyable (Q11). This result supports the users in favor of rank-based list more than the visual-based interface, but the visualization shows

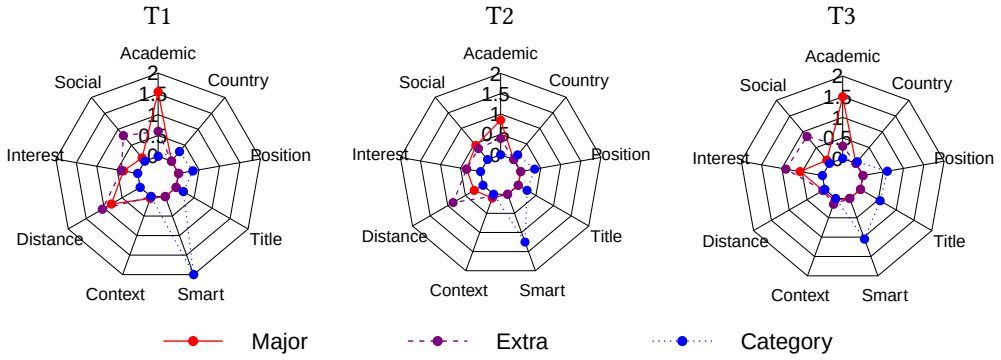


Fig. 9. The scatter plot features the usage of Study 2. The red line is the usage of major features, the purple line is the usage of extra features, and the blue line is the usage of category features.

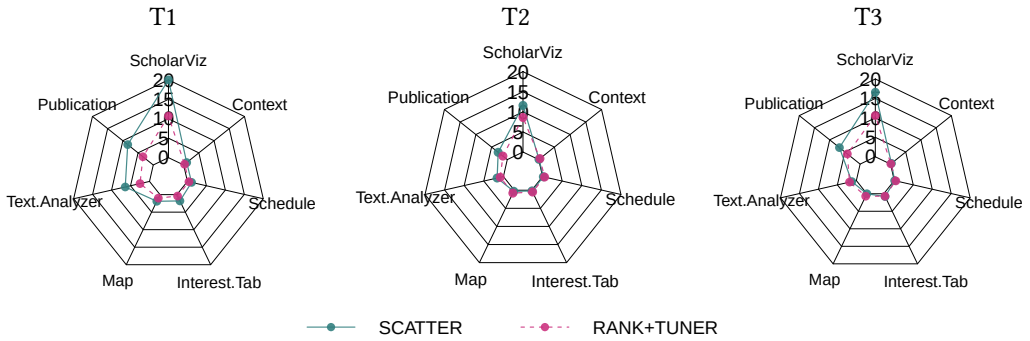


Fig. 10. The explanation function usage of Study 2. The light plum color is the usage of RANK interface, while the light blue color is the usage of SCATTER interface.

an increased level of usability on gaining trust and enjoyment in using the interface. Surprisingly, the feedback also indicates that the TUNER interface would be better for the user to fulfill the task on the feature diversity (Q1) and category diversity (Q2), but that the SCATTER interface is outperformed on the ability to Perceive Diversity (Q3).

This result shows that a user tends to use the ranked list with better controllability and transparency to conduct diversity-oriented tasks. The scatter visualization would play the role of helping the users to perceive diversity in multiple areas of relevance. The final preference result in Figure 12 also confirms this conclusion. About half of the users select the TUNER interface as the one with an advantage at helping to explore diversity, providing more informative information, being more useful, and fitting their preference - but the users also agree that the SCATTER interface could better help to perceive diversity after they finished the three tasks on two interfaces.

## 6.5 Diversity Analysis

Table 4 shows the diversity analysis of Study 2. In task 1 (relevance-oriented), there is no significant difference between the two interfaces on the diversity measurement, showing both of the interfaces can support the user to fulfill a relevance-oriented task. However, in the diversity-oriented tasks 2&3, we found the TUNER group could achieve higher entropy than the SCATTER group. This finding hints that even with a ranked-list interface the user can achieve a reasonable level of

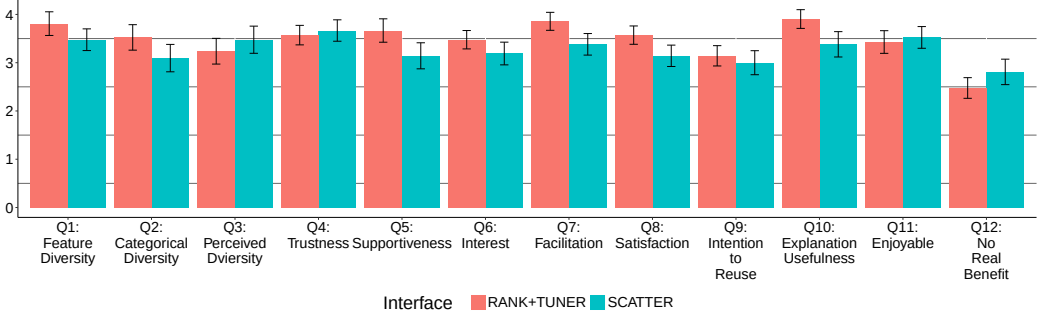


Fig. 11. User feedback analysis results for study 2: we did not find significant difference in all aspects, which indicates that the usability of two interfaces was comparable. (A cut-off value was set at 3.5 on the 5 point scale. Statistical significance level: (\*)  $p < 0.05$ .)

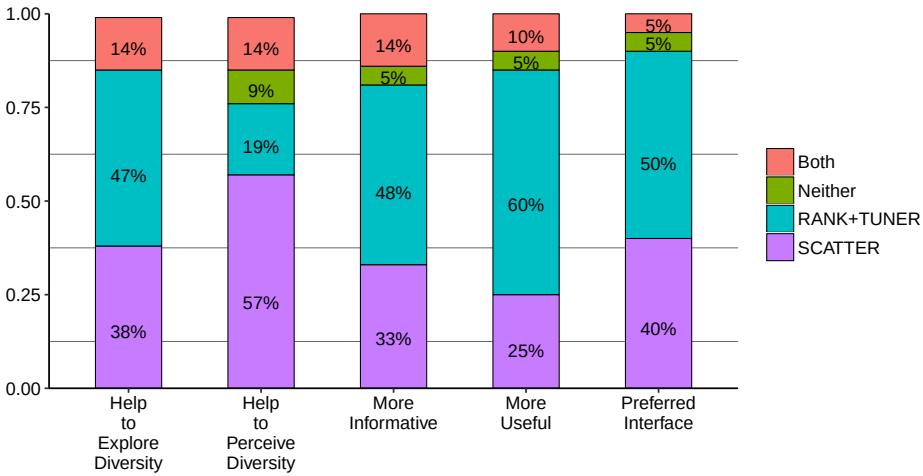


Fig. 12. User preferences for study 2 collected after the users experienced both interfaces. The result shows that the TUNER interface was preferred by users in all aspects except perceived diversity.

selection diversity if controllability and transparency for each considered dimension of relevance are available. At the same time, the SCATTER interface performed slightly (but not significantly) better than the TUNER interface in the *category* diversity metrics. This finding helps to highlight the value of color-coding data in the SCATTER interface, a feature not supported by TUNER. The diversity analysis of Social Context is omitted due to the page limitation.

## 6.6 Discussion

In Study 2, we presented a new rank-based interface for recommender attendees at a research conference. A total of five dimensions of relevance were proposed from the Personalized Relevance Model. We conducted a user study in a real conference environment to compare the two interfaces of an enhanced ranked list (TUNER) and the visualization interface (SCATTER). Our experimental results suggested the different suitable scenarios for the two interfaces. We found that, even in diversity tasks with multi-relevance settings, the users were still able to fulfill the diversity task

| Dimensions          | Task 1          |                   |   | Task 2          |                   |   | Task 3          |                   |   |
|---------------------|-----------------|-------------------|---|-----------------|-------------------|---|-----------------|-------------------|---|
|                     | TUNER<br>M (SE) | SCATTER<br>M (SE) | P | TUNER<br>M (SE) | SCATTER<br>M (SE) | P | TUNER<br>M (SE) | SCATTER<br>M (SE) | P |
| Academic + Social   | 0.12 (0.05)     | 0.12 (0.05)       |   | 0.15 (0.02)     | 0.12 (0.04)       | * | 0.16 (0.03)     | 0.13 (0.05)       |   |
| Academic + Interest | 0.17 (0.10)     | 0.16 (0.07)       |   | 0.17 (0.07)     | 0.13 (0.07)       | * | 0.14 (0.07)     | 0.14 (0.06)       |   |
| Academic + Distance | 0.13 (0.06)     | 0.12 (0.04)       |   | 0.16 (0.06)     | 0.12 (0.05)       | * | 0.17 (0.05)     | 0.14 (0.05)       | * |
| Social + Interest   | 0.23 (0.11)     | 0.23 (0.10)       |   | 0.25 (0.07)     | 0.19 (0.08)       | * | 0.32 (0.09)     | 0.26 (0.16)       |   |
| Social + Distance   | 0.23 (0.10)     | 0.22 (0.09)       |   | 0.28 (0.07)     | 0.22 (0.10)       | * | 0.32 (0.08)     | 0.26 (0.09)       | * |
| Interest + Distance | 0.24 (0.16)     | 0.29 (0.12)       |   | 0.31 (0.15)     | 0.24 (0.13)       | * | 0.28 (0.12)     | 0.26 (0.13)       |   |
| Title               | 0.12 (0.04)     | 0.16 (0.13)       |   | 0.15 (0.02)     | 0.17 (0.12)       |   | 0.15 (0.02)     | 0.14 (0.05)       |   |
| Position            | 0.29 (0.15)     | 0.27 (0.12)       |   | 0.28 (0.12)     | 0.32 (0.15)       |   | 0.16 (0.07)     | 0.22 (0.16)       |   |
| Country             | 0.35 (0.20)     | 0.57 (0.29)       | * | 0.41 (0.24)     | 0.58 (0.22)       | * | 0.48 (0.19)     | 0.54 (0.32)       |   |

Table 4. Diversity analysis for study 2: the table shows selection diversity for three tasks for each feature combination and category dimensions. The result indicates that the TUNER interface enabled users to explore a more diverse set of recommendation in diversity-oriented tasks T2 & T3. (Statistical significance level: (\*)  $p < 0.05$ .)

with a rank-based interface, but it required the support of interface controllability and transparency through visual encoding. Besides, while we found that the user would better perceive the diversity in the SCATTER interface, the user would prefer to adopt the TUNER interface to fulfill the diversity tasks.

Furthermore, when the user was interacting with the TUNER interface, the user spends more time inspecting the information on each row, instead of checking the explanation functions. This result shows that the higher level of diversity exploration was not triggered by the diversity-enhanced visualization or explanation (fewer clicks on the explanation tabs), but was instead contributed by the intention of the user reaction to the simulated diversity-oriented tasks. In the SCATTER interface, the user relies more on the explanation function and multi-relevance visualization to explore diversity-oriented tasks. Although the result showed lower entropy measurement when the user adapted to the SCATTER interface, the SCATTER interface can better help the user to perceive the diversity among multiple-relevance dimensions, based on the user feedback analysis through post-study questionnaires.

## 7 USER-CENTRIC EVALUATION

To better understand the mediation effects across the two interfaces, we conducted a structural equation model (SEM) analysis [36] to inspect the effects of the proposed interfaces on the user experiences. We used the logged data and questionnaire feedback from the two studies. There are two conditions and five summarized factors in the model (as shown in Figure 13). In objective system aspects (OSA), there are two manipulations based on the proposed interfaces. In subjective system aspects (SSA) and user experience (EXP), we proposed four factors based on the classification by [52] and our post-experiment questions. In interaction (INT), we listed the entropy of the participant's selection diversity (average of Task 2&3 that shown in Table 2 and 4). The model fit the statistics of  $\chi^2(96) = 234.68$ ,  $p < 0.01$ ,  $RMSEA = 0.18$ ,  $90\%CI : [0.152, 0.211]$ ,  $CFI = 0.941$ ,  $TLI = 0.922$ .

The model shows that the two manipulations have different positive effects on the system. The TUNER condition has a positive effect on the Perceived Usefulness and Perceived Diversity factors. It supports providing controllability and transparency on the ranked list, which helps the users

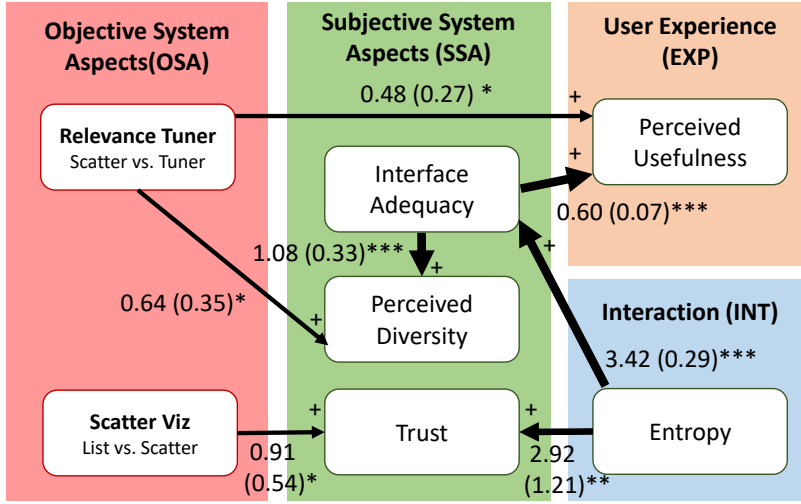


Fig. 13. The structural equation model of the experiments. The number (thickness) on the arrows represents the  $\beta$  coefficients and standard error of the effect. Significance: \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

to perceive the variety of recommendation results and increase the overall usefulness rating. The SCATTER condition only affects the Trust factor, which supports the argument for providing multi-relevance visualization to help the user to gain trust in the interface, but not mediates to the ratings of satisfaction or usefulness. We also found the effect between Interface Adequacy and Perceived Diversity, which supports a better design interface that helps the user to perceive the recommendation diversity. Furthermore, a positive impact on Interface Adequacy to Perceived Usefulness implied that the usefulness is contributed by the rating of the interface design. The Entropy value mediates the additional effect of the two interfaces. We found that the Entropy factor has a positive impact on Interface Adequacy and Trust; that is, the users who can achieve higher selection diversity through the given interfaces gain more trust and interface adequacy ratings in the system. This finding explains why the rank-based interface, with a lower learning curve, is preferred by the participants on the proposed tasks. In other words, it also supports the multi-relevance visualization that is appreciated by the users only when they are capable of taking full advantage of the interface.

## 8 USAGE ANALYSIS

The experimental systems logged interface-level actions that users perform throughout browser-based exploration such as *updating control panel*, *viewing explanation tab*, *inspect social recommendation*. To offer a deeper understanding of user work with the explored interfaces we performed a fine-grained usage analysis using a sequence mining approach. Our goal was to discover frequent usage patterns and compare them 1) between three study tasks with relevance and diversity-oriented goals and 2) between groups of the user who reported different interfaces in on post-experiment questionnaire.

### 8.1 Determining Micro Patterns for the Task-solving Behavior

We adopt the differential sequence-minding approach explained in [22]. The idea is to build compact characteristics of the problem-solving behavior of a user or a group on the level of micro-patterns.

| SCATTER |         |       |    |         |       | TUNER |         |       |    |         |       |
|---------|---------|-------|----|---------|-------|-------|---------|-------|----|---------|-------|
|         | Pattern | Supp. |    | Pattern | Supp. |       | Pattern | Supp. |    | Pattern | Supp. |
| 1       | TV      | 100%  | 16 | VQF     | 75%   | 1     | QHHH    | 100%  | 16 | QH      | 100%  |
| 2       | QF      | 95%   | 17 | ATA     | 75%   | 2     | HT      | 100%  | 17 | UV      | 100%  |
| 3       | VT      | 90%   | 18 | FY      | 75%   | 3     | VTU     | 100%  | 18 | UVT     | 95%   |
| 4       | VQ      | 90%   | 19 | TVQF    | 75%   | 4     | HTU     | 100%  | 19 | TUVT    | 95%   |
| 5       | VA      | 90%   | 20 | AV      | 75%   | 5     | HHT     | 100%  | 20 | TUT     | 95%   |
| 6       | TVQ     | 90%   | 21 | ATV     | 75%   | 6     | HHHT    | 100%  | 21 | VQ      | 90%   |
| 7       | TVA     | 90%   | 22 | AA      | 70%   | 7     | HHTU    | 100%  | 22 | VQH     | 90%   |
| 8       | TA      | 90%   | 23 | TTT     | 70%   | 8     | TUV     | 100%  | 23 | VQHH    | 90%   |
| 9       | AT      | 90%   | 24 | YT      | 70%   | 9     | QHH     | 100%  | 24 | TV      | 80%   |
| 10      | TAT     | 85%   | 25 | AVT     | 70%   | 10    | UT      | 100%  | 25 | UVTU    | 75%   |
| 11      | TF      | 85%   | 26 | FF      | 70%   | 11    | TU      | 100%  | 26 | UTV     | 75%   |
| 12      | TT      | 85%   | 27 | ATVA    | 70%   | 12    | HHHH    | 100%  | 27 | VV      | 75%   |
| 13      | TVAT    | 85%   | 28 | ATT     | 70%   | 13    | HHH     | 100%  | 28 | UTU     | 75%   |
| 14      | FAT     | 80%   | 29 | ATAT    | 65%   | 14    | VT      | 100%  | 29 | HV      | 70%   |
| 15      | FA      | 80%   | 30 | TTV     | 65%   | 15    | HH      | 100%  | 30 | VTUV    | 70%   |

Table 5. Top 30 task-solving micro patterns, ranked by support rate (the percentage of how many sequences contain the pattern). The meaning of each label is described as below. 1) Applicable in both groups: Q=starting a task; V=voting on the preferred recommendation; T=exploring an explanation tab. 2) Applicable in the SCATTER group: F=changing the control panel; Y=click on scatter plot; A=click on the ranked list. 3) Applicable in the TUNER group: H=changing the relevance slider; U=click on the ranked list.

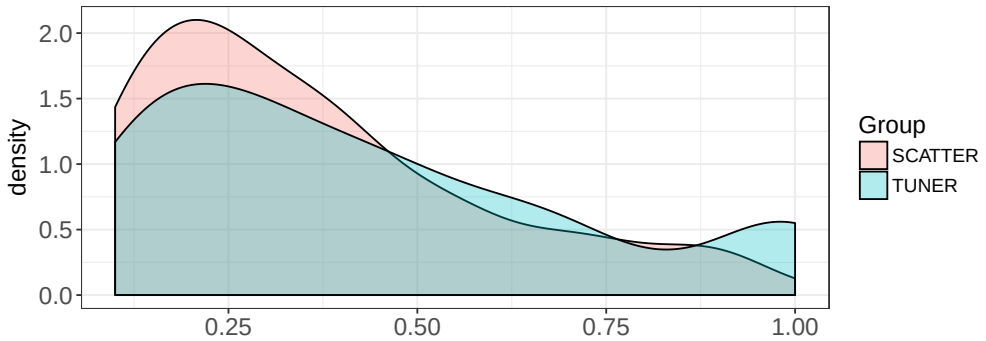


Fig. 14. Density distribution for the micro pattern support rate. Both of the distributions are right-skewed, stressing that the majority of patterns were rarely used. In the TUNER group, there are more “common” patterns that were adopted by all experiment participants. In the SCATTER group, the pattern diversity is higher.

These problem-solving patterns uncover how a user leverages application functions to solve the given search problems. In this section, we defined the “problem” as the proposed search “task” that we asked the study participants to finish. Three proposed tasks contribute to three click sequences. Our goal was to find frequent micro-patterns and then build a task-solving pattern vector to represent the frequencies of different micro-pattern occurrences in each task.

There are different click actions in the two proposed interfaces. In SCATTER, the user can make selections on the *control panel* and click on *two-dimension scatter plot*, *ranked list*, *explanation tabs* and *the voting function* (to select the preferred recommendation). In TUNER, the user can move *relevance sliders* and click on a *ranked list*, *explanation tabs* and the voting function. Each click action is assigned to a *label*. We perform the SPIRIT algorithm (Sequential Pattern Mining with Regular expression constraints) for mining frequent sequential patterns [18]. Given a sequence database  $D = s_1, s_2, \dots, s_n$ , the support (*supp*) of a pattern  $\alpha$  is the proportion of sequences of  $D$  that contains  $\alpha$  as a sub-sequence at least once. We consider a pattern to be a frequent pattern if its *support* is larger than the threshold. In this paper, we set a minimum support rate of 10% to guarantee that at least two experiment participants would repeat the same pattern. Note that the larger support rate does not indicate multiple occurrences of the pattern within a sequence since we only consider whether the pattern exists in one sequence.

We are interested in looking at pattern lengths of 2, 3 and 4. After running the whole sequence mining for combined user-task sequences, we discovered 225 patterns in the SCATTER group and 172 patterns in the TUNER group. All of the patterns were performed by at least two experiment participants ( $\alpha > 10\%$ ). The top 30 patterns and their corresponding support can be seen in Table 5. For instance, in the SCATTER group, we found “TV” pattern (click on one explanation tab and then vote for a social recommendation) happens in all sequences (*supp* = 100%). In the TUNER group, we found 17 patterns adopted by all participant while exploring the interface. The “QH HH” pattern, for example, represents the user adjusting the relevance sliders three times in a row (“H” right after the task start, as indicated by (“Q”).

We can see a clear difference between the two pattern distributions. In SCATTER, there is only one pattern supported by all participants. However, in TUNER, there are 17 patterns that occur in all sequences. This finding hints that usage patterns in SCATTER are more diverse than those in TUNER (i.e., it is used in considerably different ways by different users and in different tasks). The likely reason for that is that the SCATTER interface provides more interface components while also being less familiar to the users. To further investigate the difference, we plot the pattern density distribution of the two interfaces (Figure 14). We found that the density distribution is right-skewed, which means there are many “non-common” patterns that only occurred in a few sequences. However, there are more “common” patterns that are supported by the majority of users and fewer rare patterns in TUNER group. That is, task-solving approaches were more consistent in the TUNER interface.

## 8.2 Comparing Task-level Micro-Pattern Distributions

We analyze the task-solving pattern vectors between the three proposed tasks and across two preference groups. We used the data from study 2, which contains 20 participants (see section 6 for details). We represented participants’ system logged data as 20 sequences; each sequence consists of three task sub-sequences. Our first pattern analysis is assessing the difference between tasks. In other words, we are interested if **the problem-solving micro pattern is stable across tasks?** To understand how the assigned task impacts pattern frequencies, we analyze the pattern frequency distribution distance between the experimental tasks (sub-sequences). Following [22], we defined a pattern frequency distribution for a task or a group of users as a vector, where each frequent micro pattern (such as shown in Table 5) is represented by the frequency of its occurrence within sequences associated with a specific task or a specific group of users.

We use the Jensen-Shannon (JS) divergence to calculate the distance between pattern frequency distributions of the tasks [22]. If users approach different tasks in significantly different ways, we would expect to see that the task-level frequency distributions should significantly differ from the mixture of all tasks and each other. In contrast, if there is no task effect, then the task-level

| Pair              | SCATTER |        | TUNER  |        |
|-------------------|---------|--------|--------|--------|
|                   | M       | SE     | M      | SE     |
| ALL vs. Task 1    | 0.0390  | 0.0199 | 0.0470 | 0.0293 |
| ALL vs. Task 2    | 0.0998  | 0.0570 | 0.0590 | 0.0337 |
| ALL vs. Task 3    | 0.0596  | 0.0283 | 0.0702 | 0.0463 |
| Task 1 vs. Task 2 | 0.0525  | 0.0399 | 0.0555 | 0.0446 |
| Task 1 vs. Task 3 | 0.0549  | 0.0313 | 0.0562 | 0.0519 |
| Task 2 vs. Task 3 | 0.0608  | 0.0449 | 0.0365 | 0.0410 |

Table 6. Mean and standard error of distances between task sub-sequences and overall sequence. The vector distance between all sequences and task 1 is significantly different in the case of task 2 & 3.

patterns should not show a significant difference. The calculated distances between sequences and sub-sequences are reported in Table 6. The normality and homogeneity of the variance on pair distances are not met; thus, a non-parametric test is applied. A Mann-Whitney test is performed to test differences among the paired distances.

We found, in both groups, the task 1 sub-sequence is closer to the overall sequence. The task 2 & 3 subsequences are farther from the overall sequence in both interfaces. In other words, the patterns of user work for tasks 2 & 3 were more specific (unique) than for task 1. In the SCATTER group, we also found the distance of the task 1 sub-sequence (versus overall sequences) was significantly different than that of task 2 ( $V = 12, p < 0.001$ ) and task 3 ( $V = 43, p = 0.0192$ ). The distance between tasks 2 & 3 also shows significance ( $V = 48, p = 0.0327$ ). It means that the users approached all three tasks in considerably different ways. In the TUNER group, only the distance between task 1 and task 3 was significant ( $V = 48, p = 0.0327$ ). It shows that from the interface prospect, the users approached tasks in more similar ways. In particular, the pattern frequencies between task 2 & 3 were most similar, which could be caused by their common diversity focus or position in a sequence (more experience gained).

The result indicates that the nature of the task considerably impacts the task-solving micro-patterns. There are two possible speculations about this finding. First, the effect of experience: in task 1, the experiment participants usually took time to get familiar with the interface. Hence, the pattern of the first task might be different than the other two tasks, simply because of its position in the sequence. Second, task difference: task 1 is *relevance-oriented* and tasks 2 & 3 are *diversity-oriented*. The task effect might cause the users to adopt a different behavioral pattern in fulfilling the tasks. The finding encourages further analysis of micro-behavior patterns - in exploring the user behavior patterns to determine where the differences between tasks and user groups really occur.

### 8.3 Exploring the Task-solving Micro Patterns

In this section, we plan to compare frequently found patterns between tasks. The goal is to find pattern level differences between tasks. The top 50 micro-patterns of the SCATTER and TUNER interfaces, split by tasks, are shown in Figure 15 and 18, respectively. We also want to explore the extent to which frequent patterns exhibited by a user might be determined by the individual differences, rather than the nature of the task. To assess this hypothesis, we divided the participants into two clusters, based on their post-experiment preference of the two interfaces. As shown in Figure 12 (the question of “Preferred Interface”), users were almost equally divided into two large groups, with one preferring SCATTER and one preferring TUNER. We considered user preference to be the most valuable registered between-user difference that could affect user behavior patterns.



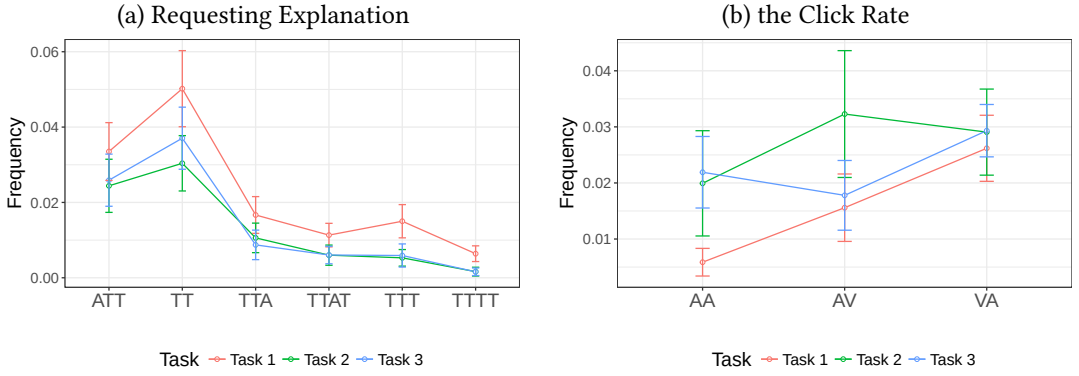


Fig. 16. (SCATTER) Highlighted task-solving micro patterns between tasks. Figure (a) shows the patterns of checking on explanation functions, while Figure (b) shows the action of clicking on the ranked list.

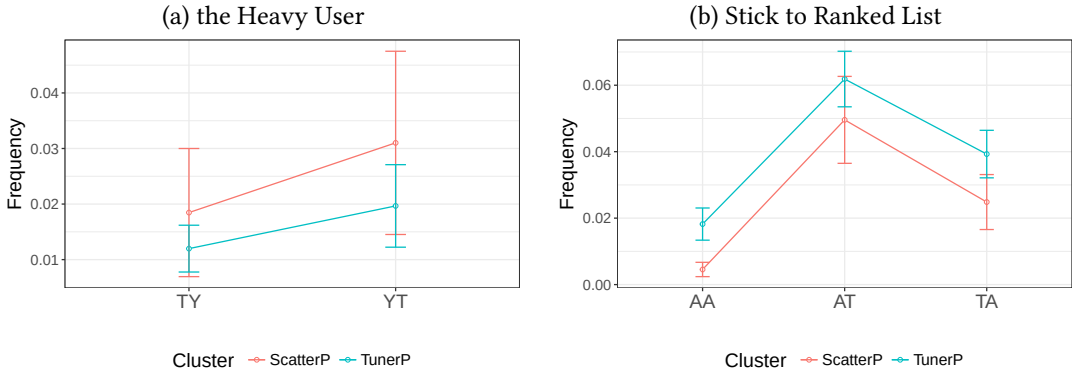


Fig. 17. (SCATTER) Highlighted task-solving micro patterns between clusters. Figure (a) shows the action of clicking on scatter plot, while Figure (b) shows the action of clicking on the ranked list and explanations.

leverage its functions due to their abilities. These participants used the scatter plot more heavily, and at the end, reported higher satisfaction with this interface. If this was not the case, then a ranked list was adopted to fulfill the tasks, regardless of the task's nature.

#### • Pattern Analysis: TUNER

Figure 18 shows micro-pattern distribution across the three tasks solved through the TUNER interface. As the preliminary distance analysis demonstrated, the pattern distribution of task 1 is different from the other two tasks. We found that the most visible difference is in changing the relevance sliders; e.g., the “HH”, “HHH” and “HHHH” patterns, which are the dominant pattern in all three tasks. A higher tuning rate was identified in the first task (shown in Figure 19(a)), which means that the users re-tune the social recommendation more in the beginning phase, likely during the process of getting familiar with the interface. On the other hand, we found that the users requested more explanations in tasks 2 & 3. For example, the “TT”, “TUV”, “VTU”, “TUT” and “UTU” patterns occur more frequently in the diversity-oriented tasks. The other interesting finding is a substantial use of “sequential voting” patterns in tasks 2 & 3, i.e., “VV” and “UVV” and “VVT”



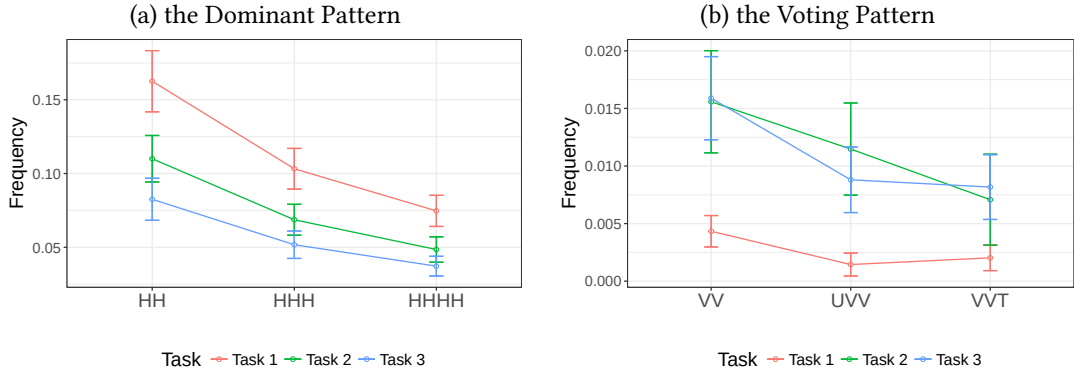


Fig. 19. (TUNER) Highlighted task-solving micro-patterns between tasks. Figure (a) shows the hybrid patterns of relevance slider re-tuning. Figure (b) shows the action of voting on the social recommendations.

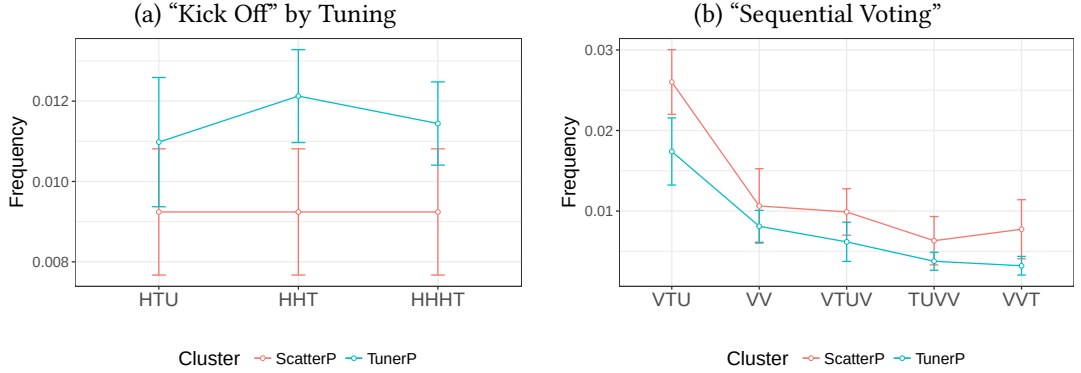


Fig. 20. (TUNER) Highlighted task-solving micro patterns between clusters. Figure (a) shows the relevance tuning was an essential action, while Figure (b) shows the action of clicking on the ranked list.

## 9 CONCLUSION

In this paper, we presented experiments with three different social recommender interfaces: Ranked List (RANK), *Scatter Viz* (SCATTER) and *Relevance Tuner* (TUNER). We first showed that providing a scatter plot (SCATTER) can help the user to better fulfill the diversity-related tasks, as compared to a simple ranked list (RANK). However, despite the benefits of the new two-dimensional presentation, the users still extensively used the ranked list component of the interface.

Based on the results of the first study, we attempted to integrate the ability to coordinate multiple aspects of relevance within the ranked list rather than offering it in a separate component as in SCATTER. To compensate for the biasing nature of the ranked list, we also provided a controllable fusion of relevance aspects. The resulting expanded ranked list interface (TUNER) offered both controllability and visual encoding of multiple relevance aspects. We showed that the users could adopt a rank-based list to fulfill diversity-oriented tasks with higher selection diversity. The usability analysis revealed that both SCATTER and TUNER were ranked by the conference users with high subjective ratings. However, TUNER required less learning effort. We also discussed the mediation

effects of the proposed interfaces on the user experience. The analysis helps to describe the benefits of the two proposed interfaces in social recommender systems.

One of our goals was to understand how users apply the recommender interfaces to diversity-oriented tasks. We prepared tasks that encourage users to explore the conference attendees using multiple aspects of relevance. We found that both of the proposed interfaces were capable of helping the user to fulfill the assigned tasks. The experimental result supported the finding that the participants were able to correlate multiple aspects of relevance using two dimensions of visualization in SCATTER and the controllable ranked list with multi-aspect visualization in TUNER. We found that an extension of a traditional ranked list with controllability and visualization was better than a separate visualization component in the sense of on getting familiar with the new interface (which led to a higher rating on the user preference). In contrast, a separate diversity-enhanced visualization can also achieve the goal, but it came at the cost of a steeper learning curve. However, once the users were familiar with the interface, it brought an advantage of helping the users to perceive diversity and gain trust in the recommendations.

A sequential analysis of user actions provides a deeper understanding of the user interactions with the two proposed interfaces. We extracted a set of frequently occurring “task-solving micro-patterns” for an in-depth user behavior analysis. We saw significant differences on the micro-pattern level between all tasks while also observing that tasks 2 & 3 were more similar to each other than to task 1. We believe that prior experience might give an advantage to interfaces based on a ranked list, as most users have been familiar with a ranked-list presentation through Web search and other popular interfaces. A user might benefit from a range of additional features such as a two-dimensional visual interface, controllability, and explainability; however, the mere availability of these components does not mean that the user will adopt them to fulfill their information needs. The learning effort should be considered while deploying the interface to real-world applications. A lesson we have learned from this project is: *the users leverage the diversity-enhanced interface design only when they are prepared to use it (by their abilities or experience)*. A simpler approach might be ultimately appreciated by a much larger number of users setting up a challenge to HCI researchers to balance the functionality and usability in interfaces that are focused on enhancing social recommendation diversity.

These findings shed light on designing diversity-enhanced and diversity-aware interfaces in a social recommender system. A *diversity-enhanced* interface could offer extra help to users working with a diversity-oriented task. For example, a conference attendee exploring social recommendations within a particular conference might need to find a group of users satisfying specific diversity criteria. However, users may not specifically think about diversity when interacting with a recommender system. For example, when accessing a news article on social media, a user may not have an intention to read news articles from a different perspective. In this case, we need a *diversity-aware* interface to help the user to perceive the prospective diversity, which might help to solve the filter-bubble effect [59]. In this paper, we find that the effects of usability and perceived diversity are additive. The best user experience may happen when the users can take advantages of both types of interfaces. It remains an open question as to how to design an interface to reflect the needs of users better.

Finally, it could be observed that the core problem explored in this paper, i.e., diversity-promoting interfaces for the multi-aspect recommendation, is more general than the context in which we explored it, i.e., recommending people at academic conferences. Our diversity-promoting interfaces can be implemented to recommend social media content, research paper, news, movies and other kinds of items with multiple relevance aspects or similarity features. To avoid over-generalization of our findings, this paper specifically focused on the problem of people recommendation supported by our studies. In our future works, however, we plan to examine the generalizability of our

proposed visual diversity-promoting interfaces and observe the user exploration patterns in different exploratory search tasks.

## 10 LIMITATIONS AND FUTURE WORK

Our study has some limitations. First, the within-subject user study was conducted using consecutive years of the same conference series. Some well-known and senior domain experts may appear in the recommendation list for both conferences. This repetition may cause bias in our user studies. Second, the data sparsity and cold-start problem may hurt the recommendation performance; for example, the Interest feature is less useful for users who never bookmarked any talks or papers within the Conference Navigator System (CN3). We tried our best to send out emails both before and during the conference to improve interest-based recommendations. Third, the scale of reported user studies is relatively small. It may decrease the statistical power of the findings. Fourth, the experiment was conducted at mid-size conferences, so we were not able to explore scaling issues which might occur at conferences with a much larger number of attendees or in a different recommendation context with a large set of items to explore. We hope to address some of these limitations in our future work.

## A QUESTIONNAIRE USED IN STUDIES

### A.1 Post-Stage Questionnaire

|     |                                                                                                                         |
|-----|-------------------------------------------------------------------------------------------------------------------------|
| Q1  | The interface helps me to explore various interesting people in the conference.                                         |
| Q2  | It is helpful to see people attributes like Title, Country, and Position when exploring interesting people in the list. |
| Q3  | The interface helps me to perceive the diversity of explored attendees                                                  |
| Q4  | The interface helps me to improve my trust in the people recommendation result.                                         |
| Q5  | The interface helps me to understand why specific attendees were recommended.                                           |
| Q6  | I like the people recommendation result from the system.                                                                |
| Q7  | I became familiar with the system very quickly.                                                                         |
| Q8  | Overall, I am satisfied with the system.                                                                                |
| Q9  | I will frequently use the system at future conferences.                                                                 |
| Q10 | It was useful to see the explanation of scores produced by different recommendation components.                         |
| Q11 | It is fun to use the system.                                                                                            |
| Q12 | The system has no real benefit for me.                                                                                  |

Table 7. The questions of Post-Stage Questionnaire [8, 52].

### A.2 Post-Experiment Questionnaire

|    |                                                                                                                                                |
|----|------------------------------------------------------------------------------------------------------------------------------------------------|
| Q1 | Which recommendation interface did you prefer?                                                                                                 |
| Q2 | Which recommendation interface did you find more informative?                                                                                  |
| Q3 | Which recommendation interface did you find more useful?                                                                                       |
| Q4 | Which recommendation interface was better at helping you to perceive the diversity of recommendations?                                         |
| Q5 | Which recommendation interface was better at helping you to explore the diversity of recommendation through different features and categories? |

Table 8. The questions of Post-Experiment Questionnaire [29].

## REFERENCES

- [1] Jisun An, Daniele Quercia, and Jon Crowcroft. 2013. Why individuals seek diverse opinions (or why they don't). In *Proceedings of the 5th Annual ACM Web Science Conference*. ACM, 15–18.
- [2] Fabiano M Belém, Carolina S Batista, Rodrygo LT Santos, Jussara M Almeida, and Marcos A Gonçalves. 2016. Beyond Relevance: Explicitly Promoting Novelty and Diversity in Tag Recommendation. *ACM Transactions on Intelligent Systems and Technology (TIST)* 7, 3 (2016), 26.
- [3] Toine Bogers. 2018. Tag-based recommendation. In *Social Information Access*. Springer, 441–479.
- [4] Dirk Bollen, Bart P Knijnenburg, Martijn C Willemsen, and Mark Graus. 2010. Understanding choice overload in recommender systems. In *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 63–70.
- [5] Svetlin Bostandjiev, John O'Donovan, and Tobias Höllerer. 2012. TasteWeights: a visual interactive hybrid recommender system. In *Proceedings of the sixth ACM conference on Recommender systems*. ACM, 35–42.
- [6] Engin Bozdag and Jeroen van den Hoven. 2015. Breaking the filter bubble: democracy and design. *Ethics and Information Technology* 17, 4 (2015), 249–265.
- [7] Peter Brusilovsky, Jung Sun Oh, Claudia López, Denis Parra, and Wei Jeng. 2016. Linking information and people in a social system for academic conferences. *New Review of Hypermedia and Multimedia* (2016), 1–31.
- [8] Bruno Cardoso, Gayane Sedrakyan, Francisco Gutiérrez, Denis Parra, Peter Brusilovsky, and Katrien Verbert. 2018. IntersectionExplorer, a Multi-Perspective Approach for Exploring Recommendations. *International Journal of Human-Computer Studies* (2018).
- [9] Jilin Chen, Werner Geyer, Casey Dugan, Michael Muller, and Ido Guy. 2009. Make new friends, but keep the old: recommending people on social networking sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 201–210.
- [10] Henriette Cramer, Vanessa Evers, Satyan Ramalal, Maarten Van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. 2008. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction* 18, 5 (2008), 455.
- [11] Paolo Cremonesi, Franca Garzotto, and Roberto Turrin. 2012. Investigating the persuasion potential of recommender systems from a quality perspective: An empirical study. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 2 (2012), 11.
- [12] Tuan Nhon Dang and Leland Wilkinson. 2014. Scagexplorer: Exploring scatterplots by their scagnostics. In *Visualization Symposium (PacificVis), 2014 IEEE Pacific*. IEEE, 73–80.
- [13] Stavín Deeswe and Raymond Kosala. 2015. An integrated search interface with 3D visualization. *Procedia Computer Science* 59 (2015), 483–492.
- [14] Cecilia di Sciascio, Vedran Sabol, and Eduardo E Veas. 2016. Rank As You Go: User-Driven Exploration of Search Results. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 118–129.
- [15] Michael D Ekstrand, F Maxwell Harper, Martijn C Willemsen, and Joseph A Konstan. 2014. User perception of differences in recommender algorithms. In *Proceedings of the 8th ACM Conference on Recommender systems*. ACM, 161–168.
- [16] Michael D Ekstrand, Daniel Kluver, F Maxwell Harper, and Joseph A Konstan. 2015. Letting users choose recommender algorithms: An experimental study. In *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, 11–18.
- [17] Siamak Faridani, Ephrat Bitton, Kimiko Ryokai, and Ken Goldberg. 2010. Opinion space: a scalable tool for browsing online comments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1175–1184.
- [18] Minos N Garofalakis, Rajeev Rastogi, and Kyuseok Shim. 1999. SPIRIT: Sequential pattern mining with regular expression constraints. In *VLDB*, Vol. 99. 7–10.
- [19] Mouzhi Ge, Carla Delgado-Battenfeld, and Dietmar Jannach. 2010. Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 257–260.
- [20] Dorota Glowacka, Tuukka Ruotsalo, Ksenia Konuyshkova, Samuel Kaski, Giulio Jacucci, et al. 2013. Directing exploratory search: Reinforcement learning from user interactions with keywords. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 117–128.
- [21] Brynjar Gretarsson, John O'Donovan, Svetlin Bostandjiev, Christopher Hall, and Tobias Höllerer. 2010. Smallworlds: visualizing social recommendations. In *Computer Graphics Forum*, Vol. 29. Wiley Online Library, 833–842.
- [22] Julio Guerra, Shaghayegh Sahebi, Yu-Ru Lin, and Peter Brusilovsky. 2014. The problem solving genome: Analyzing sequential patterns of student work with parameterized exercises. (2014).
- [23] Pankaj Gupta, Ashish Goel, Jimmy Lin, Aneesh Sharma, Dong Wang, and Reza Zadeh. 2013. Wtf: The who to follow service at twitter. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 505–514.
- [24] Ido Guy. 2015. Social recommender systems. In *Recommender Systems Handbook*. Springer, 511–543.
- [25] Ido Guy. 2018. *People Recommendation on Social Media*. Springer International Publishing, Cham, 570–623. [https://doi.org/10.1007/978-3-319-90092-6\\_15](https://doi.org/10.1007/978-3-319-90092-6_15)

- [26] Ido Guy, Inbal Ronen, and Eric Wilcox. 2009. Do you know?: recommending people to invite into your social network. In *Proceedings of the 14th international conference on Intelligent user interfaces*. ACM, 77–86.
- [27] Ido Guy, Sigalit Ur, Inbal Ronen, Adam Perer, and Michal Jacovi. 2011. Do you want to know?: recommending strangers in the enterprise. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, 285–294.
- [28] Jonathan L Herlocker, Joseph A Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. ACM, 241–250.
- [29] Rong Hu and Pearl Pu. 2011. Helping Users Perceive Recommendation Diversity.. In *DiveRS@ RecSys*. 43–50.
- [30] Marius Kaminskas and Derek Bridge. 2016. Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems. *ACM Trans. Interact. Intell. Syst.* 7, 1, Article 2 (Dec. 2016), 42 pages. <https://doi.org/10.1145/2926720>
- [31] Marius Kaminskas and Derek Bridge. 2016. Diversity, Serendipity, Novelty, and Coverage: A Survey and Empirical Analysis of Beyond-Accuracy Objectives in Recommender Systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 7, 1 (2016), 2.
- [32] Hannah Kim, Jaegul Choo, Haesun Park, and Alex Endert. 2016. Interaxis: Steering scatterplot axes via observation-level interaction. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 131–140.
- [33] Khalil Klouche, Tuukka Ruotsalo, Diogo Cabral, Salvatore Andolina, Andrea Bellucci, and Giulio Jacucci. 2015. Designing for exploratory search on touch devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 4189–4198.
- [34] Khalil Klouche, Tuukka Ruotsalo, Luana Micallef, Salvatore Andolina, and Giulio Jacucci. 2017. Visual Re-Ranking for Multi-Aspect Information Retrieval. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. ACM, 57–66.
- [35] Daniel Kluver, Michael D Ekstrand, and Joseph A Konstan. 2018. Rating-based collaborative filtering: algorithms and evaluation. In *Social Information Access*. Springer, 344–390.
- [36] Bart P. Knijnenburg, Svetlin Bostandjiev, John O'Donovan, and Alfred Kobsa. 2012. Inspectability and Control in Social Recommenders. In *6th ACM Conference on Recommender System*. 43–50.
- [37] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. 2017. User Preferences for Hybrid Explanations. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*. ACM, 84–88.
- [38] Johannes Kunkel, Benedikt Loepp, and Jürgen Ziegler. 2017. A 3D item space visualization for presenting and manipulating user preferences in collaborative filtering. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces*. ACM, 3–15.
- [39] Danielle Lee and Peter Brusilovsky. 2018. Recommendations based on social links. In *Social Information Access*. Springer, 391–440.
- [40] Q Vera Liao and Wai-Tat Fu. 2013. Beyond the filter bubble: interactive effects of perceived threat and topic involvement on selective exposure to information. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2359–2368.
- [41] Sean M McNee, John Riedl, and Joseph A Konstan. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems*. ACM, 1097–1101.
- [42] Jennifer Moody and David H Glass. 2016. A Novel Classification Framework for Evaluating Individual and Aggregate Diversity in Top-N Recommendations. *ACM Transactions on Intelligent Systems and Technology (TIST)* 7, 3 (2016), 42.
- [43] Sean A Munson and Paul Resnick. 2010. Presenting diverse political opinions: how and how much. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1457–1466.
- [44] Mark EJ Newman. 2001. Clustering and preferential attachment in growing networks. *Physical Review E* 64, 2 (2001), 025102.
- [45] Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web*. ACM, 677–686.
- [46] John O'Donovan, Brynjar Gretarsson, Svetlin Bostandjiev, Tobias Hollerer, and Barry Smyth. 2009. A visual interface for social information filtering. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, Vol. 4. IEEE, 74–81.
- [47] John O'Donovan, Barry Smyth, Brynjar Gretarsson, Svetlin Bostandjiev, and Tobias Höllerer. 2008. PeerChooser: visual interactive recommendation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1085–1088.
- [48] Michael P O'Mahony and Barry Smyth. 2018. From opinions to recommendations. In *Social Information Access*. Springer, 480–509.
- [49] Valeria Orso, Tuukka Ruotsalo, Jukka Leino, Luciano Gamberini, and Giulio Jacucci. 2017. Overlaying social information: The effects on users' search and information-selection behavior. *Information Processing & Management* 53, 6 (2017), 1269–1286.

- [50] Anshul Vikram Pandey, Josua Krause, Cristian Felix, Jeremy Boy, and Enrico Bertini. 2016. Towards understanding human similarity perception in the analysis of large sets of scatter plots. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 3659–3669.
- [51] Denis Parra and Peter Brusilovsky. 2015. User-controllable personalization: A case study with SetFusion. *International Journal of Human-Computer Studies* 78 (2015), 43–67.
- [52] Pearl Pu, Li Chen, and Rong Hu. 2011. A user-centric evaluation framework for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*. ACM, 157–164.
- [53] Tuukka Ruotsalo, Jaakko Peltonen, Manuel Eugster, Dorota Glowacka, Ksenia Konyushkova, Kumaripaba Athukorala, Ilkka Kosunen, Aki Reijonen, Petri Myllymäki, Giulio Jacucci, et al. 2013. Directing exploratory search with interactive intent modeling. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. ACM, 1759–1764.
- [54] J Ben Schafer, Joseph A Konstan, and John Riedl. 2002. Meta-recommendation systems: user-controlled integration of diverse recommendations. In *Proceedings of the eleventh international conference on Information and knowledge management*. ACM, 43–51.
- [55] Rory L.L. Sie, Hendrik Drachler, Marlies Bitter-Rijpkema, and Peter Sloep. 2012. To whom and why should I connect? Co-author recommendation based on powerful and similar peers. *Int. J. Technology Enhanced Learning* 4, 1/2 (2012), 121–137.
- [56] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 990–998.
- [57] Robert Tarjan. 1972. Depth-first search and linear graph algorithms. *SIAM journal on computing* 1, 2 (1972), 146–160.
- [58] Choon Hui Teo, Houssam Nassif, Daniel Hill, Sriram Srinivasan, Mitchell Goodman, Vijai Mohan, and SVN Vishwanathan. 2016. Adaptive, Personalized Diversity for Visual Discovery. In *Proceedings of the 10th ACM Conference on Recommender Systems*. ACM, 35–38.
- [59] Nina Tintarev. 2017. Presenting Diversity Aware Recommendations: Making Challenging News Acceptable. (2017).
- [60] Nava Tintarev and Judith Masthoff. 2012. Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction* 22, 4-5 (1 Oct. 2012), 399–439.
- [61] Chun-Hua Tsai. 2017. An Interactive and Interpretable Interface for Diversity in Recommender Systems. In *Proceedings of the 22Nd International Conference on Intelligent User Interfaces Companion (IUI '17 Companion)*. ACM, New York, NY, USA, 225–228.
- [62] Chun-Hua Tsai and Peter Brusilovsky. 2016. A personalized people recommender system using global search approach. *IConference 2016 Proceedings* (2016).
- [63] Chun-Hua Tsai and Peter Brusilovsky. 2017. Enhancing Recommendation Diversity Through a Dual Recommendation Interface. In *Workshop on Interfaces and Human Decision Making for Recommender Systems*.
- [64] Chun-Hua Tsai and Peter Brusilovsky. 2017. Leveraging Interfaces to Improve Recommendation Diversity. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*. ACM, 65–70.
- [65] Chun-Hua Tsai and Peter Brusilovsky. 2017. Providing Control and Transparency in a Social Recommender System for Academic Conferences. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*. ACM, 313–317.
- [66] Chun-Hua Tsai and Yu-Ru Lin. 2016. Tracing and Predicting Collaboration for Junior Scholars. In *Proceedings of the 25th International Conference Companion on World Wide Web*. International World Wide Web Conferences Steering Committee, 375–380.
- [67] Katrien Verbert, Denis Parra, Peter Brusilovsky, and Erik Duval. 2013. Visualizing recommendations to support exploration, transparency and controllability. In *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 351–362.
- [68] Jesse Vig, Shilad Sen, and John Riedl. 2012. The tag genome: Encoding community knowledge to support novel interaction. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 3 (2012), 13.
- [69] Wei-quan Wang and Izak Benbasat. 2007. Recommendation agents for electronic commerce: Effects of explanation facilities on trusting beliefs. *Journal of Management Information Systems* 23, 4 (2007), 217–246.
- [70] David Wong, Siamak Faridani, Ephrat Bitton, Björn Hartmann, and Ken Goldberg. 2011. The diversity donut: enabling participant control over the diversity of recommended responses. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1471–1476.
- [71] Bo Xiao and Izak Benbasat. 2007. E-commerce product recommendation agents: use, characteristics, and impact. *MIS quarterly* 31, 1 (2007), 137–209.

Received February 2007; revised March 2009; accepted June 2009